A mechanistic insight on the role
of visual-attention in context-dependent preference reversal effects

Dissertation submitted to the
Faculty of
Business, Economics and Informatics
of the University of Zurich

to obtain the degree of
Doktorin der Neuroökonomie, Dr. sc.
(corresponds to Doctor of Neuroeconomics, PhD)

presented by

Gaia Lombardi
from Italy

approved in July 2020 at the request of
Prof. Dr. Ernst Fehr
Prof. Dr. Todd Hare
Prof. Dr. Rafael Polania

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, 15.07.2020

Chairman of the Doctoral Board: Prof. Dr. Todd Hare

Acknowledgments

This section of my dissertation will be in all probability one of the most read chapter among all. Thus, I would like to take the time and the opportunity to say few important words about my research papers. They all are in preparation, however, I believe they are quite cool. Most importantly, they can be found in the Appendix to this dissertation, so go and read them!

Alright, I also would like to express my deep gratitude to all the people who made this work possible, and without whom I could have not written the cool papers I was mentioning above. Thus, thank you people!

Table D.1 in the Appendix shows the complete list of the people I would like to thank [1].

[1] For any complaint about missing names please feel free to contact gaia.lombardi@econ.uzh.ch

Abstract

A mechanistic insight on the role of visual-attention in context-dependent preference reversal effects

Classical economic theories of rational choices assume that individuals have well-defined and ordered preferences which are revealed at the time of the choice. However, experimental evidence has shown that these assumptions do not always hold, and preferences can be affected by the context and by the choice set in which the decision is made. Notable violations of these principles are the framing and the decoy effects, which are clear examples of preference reversals in context-dependent decision making.

Despite a lack of deep understanding of the inner mechanisms or a consensus on the driven factors that generate these preference reversals, many computational models have been proposed in the cognitive science literature to account for these types of context-dependent effects. However, none of the theories is able to explain all of the effects at the same time, and the models have to be adapted for the different contexts to account for the different preference reversals. Thus, a still open question in cognitive sciences is whether there exists one inner mechanism that is related to all of the different context-dependent reversals.

Here, I investigate the hypothesis that changes in allocation of attention driven by the changing context could be a candidate common mechanism for preference reversals in decision-making. The intuition for this prediction comes from a growing stream of evidence in decision neuroscience showing a high impact of attention on choice frequency, and from the observation that in most of the context-dependent decisions where these context-dependent effects are generated what changes in the decision set is only the focus of the decision makers on different aspects of the decision.

Thus, I combine behavioral paradigms, eye-tracking and computational techniques to examine the role of visual-attention as a proxy for attention in context-dependent changes of preferences during risky decision problems.

Three experimental and computational studies are reported in this dissertation. In the first study, we investigate the framing effect based on an attentional drift diffusion model (aDDM) framework, and find evidence for a possible decision process in which shifts of attention allocation are the main source of changes in choice frequency without directly altering the valuation of the options in the choice set. In the second study, the decoy effect is examined and the computational predictions of existing models are tested in our data. We find attentional related effects on choice frequency which have not been recorded before, and we explore whether there are existing models of attention that can account for our finding which predictions can be found in the empirical data.

In this dissertation, I also discuss the importance of model fitting and recovery parameters analysis when dealing with computational models in cognitive science, and I propose a standard procedure to follow when dealing with new data sets or new models. In the third study, we develop a method to quickly and efficiently fit hierarchical Bayesian drift diffusion models that have piecewise constant drift rates, and we discuss in details the recovery parameters procedure prior to the empirical model fitting. This method is used in the first study and aim to improve the quality and the accuracy of the model fitting.

Table of Contents

List Of Manuscripts

The dissertation is based on the following research articles:

Study 1

Lombardi G, Hare TA, Fehr E. A mechanistic foundation of the role of visual-attention in the framing effect. *In preparation*

Study 2

Lombardi G, Hare TA, Fehr E. Is the decoy effect an attention-driven phenomenon? *In preparation*

Study 3

Lombardi G, Hare TA. Simple method to fit hierarchical Bayesian piecewise time-constant drift diffusion models. *In preparation*

Chapter 1: Introduction

Many classical economic theories of decision making are based of the assumption that choices are revealed by a well-defined preference order for each individual over any set of options. Thus, each option has some personal value (or utility) for the decision maker and a well established position in her preference ordering which does not depend on the context or on the choice set of the decision. A practical implication of this principle can be illustrated with the following example. If, for instance, the reader prefers spending her free time reading this dissertation on *A mechanistic insight on the role of visual-attention in context-dependent preference reversal effects* versus Jules Verne's adventure book *Twenty Thousand Leagues Under The Sea*, she does not prefer spending her free time reading Jules Verne's book instead of my dissertation. This concept seems to be simple and intuitive at a first glance, however, when the choice is not as obvious as preferring this dissertation over an adventure novel, the assumption does not always hold and experimental evidence has suggested that preferences can be affected by the context and by the choice set in which the decision is made. Decision makers do not appear to have clear preferences before the choice is faced and do not seem to know a priori what they prefer before encountering the options in a specific choice context. For example, let's assume that before facing the above decision the reader is falsely framed about how boring this dissertation is going to be. Then, she might want to choose reading Verne's book now, even though the two options in the choice set were kept exactly the same. Psychology and cognitive sciences research have widely explored this and other type of framing effects (Kahneman and Tversky,1979; Tversky and Kahneman, 1981; Shafir 1993), but very few computational models of decision making in Economics can account for it (Kahneman and Tversky, 1979).

Another example of a violation of this assumption is the decoy effect. Suppose again that our reader is offered the choice between this manuscript and *Twenty Thousand Leagues*

Under The Sea, but now she vacillates between these two options until she convinces her self that reading this dissertation is the right way to spend her free time. However, just before she can start reading all about visual-attention in preference reversals effects, she realizes that there is another option on the book shelf, she could also use her time reading *The Mighty Orinoco* of which author is once more Jules Verne. This changes everything, now she knows she definitely wants to read *Twenty Thousand Leagues Under The Sea.*

Several examples of such paradoxical reversal behavior have been recorded in many studies in a wide ranges of contexts and across species since the earliest 70s (Tversky, 1972; Heath and Chatterjee, 1995; Shafir et al., 2002; Trueblood et al., 2013; Berkowitsch et al., 2014). All of the different decoy effects (such as attraction effect, similarity effect, compromise effect, etc.) violate the independence of irrelevant alternative (IIA) principle which is a fundamental property of the majority of choice theories such as expected utility theory (Von Neumann and Morgenstern, 1947) and is implied by common choice rules such as Luce's choice axioms or the softmax choice rule (Luce, 1959; Sutton and Barto, 1998).

Despite a lack of deep understanding of the inner mechanisms or a consensus on the driven factors that generate these decoy effects, many computational models have been proposed to account for these types of preference reversal (Kahneman and Tversky, 1984; Busemeyer and Townsend, 1993; Hotaling et al., 2010; Roe et al., 2001; Usher and McClelland, 2004; Bhatia, 2013; Noguchi and Stewart 2018; Gluth et al. 2018).

Most of the computational modelling in decision-making over the past 25 years have focused on integrating models from perceptual processing into value-based (or preference-based) decision-making. Notable examples of this attempt are sequential sampling models which have become the dominant theory in the cognitive sciences and have also started to play a central role in decision neuroscience (Ratcliff et al. 2016; Forstmann et al. 2016; Hanks and Summerfield 2017). The rise of sequential sampling models can be attributed to mainly three reasons: 1) they have been shown to be good at fitting a wide variety of experimental data set (Ractliff and Smith 2004; Ractliff et al. 2016), 2) they can simultaneously account for choices and response time distributions – i.e. trade-off speed accuracy, and 3) there is a growing body of evidence suggesting that the firing properties of neurons that likely drive decisions in the Lateral intraparietal cortex (LIP) and the frontal eye field (FEF) areas are well described by stochastic sequential sampling models (e.g., Ditterich, 2006; Gold and Shadlen 2007; Churchland et al., 2008; Purcell et al. 2012).

In order to reproduce context-dependent effects with computational modelling, different features have been added to sequential sampling models. Some research pointed out the need of multi-attributes models for decision making (Busemeyer et al. 2019), others lateral inhibition leakage (Roe et al., 2001; Usher and McClelland, 2004), or loss aversion features (Usher and McClelland, 2004) and saliency (Tsetsos and Usher 2012; Towal et al. 2013). However, all of these added features cannot explain more than some context-dependent effects on choices at the same time, and the models have to be adapted for the different contexts to account for preference reversals. Thus, a still open question in cognitive sciences is whether there exists one inner mechanism that is related to all of the different context-dependent effects. One main aim of this thesis is to test whether attention could be a candidate common mechanism for preference reversals in context-dependent decision-making.

The intuition for this hypothesis comes from the observation that in most of the context-dependent decisions where changes in preferences are observed what actually changes in the decision set is the focus by the decision maker on different aspects of the decision. For instance, in our reader example above, introducing a third option which is very similar to one of the other options, i.e. *The Mighty Orinoco* by Jules Verne is similar to the second option *Twenty Thousand Leagues Under The Sea* by the same author, put the similar original option on a different light. In our example, we could speculate that our reader's willingness to read an adventure book by Jules Verne is reinforced by the introduction of the third option. Now, reading a science fiction book attracts more attention, and in turns more weight than reading a long dissertation on the role of visual-attention in preference reversal effects.

Further less speculative reasons for believing that attention could be an inner mechanism which can affect choices and drive preference reversals can be found in the recent growing literature about attention in decision-making. This stream of research has shown a high impact of attention on choice frequency, namely the decision-makers are more likely to choose options they have attended longer (Krajbich et al. 2010; Krajbich and Rangel 2011; Krajbich and Rangel 2012; Towal et al. 2013), and that this bias is present even when the attention is experimentally manipulated (Shimojo et al. 2003; Armel et al. 2008; Milosavljevic et al., 2010; Tavares et al. 2017). Second, attention has previously been linked to preference reversals in the decoy effect (Noguchi and Stewart 2018) and some models have incorporated it in their framework (Krajbich and Rangel, 2011; Roe et al., 2001; Usher and McClelland, 2004; Bhatia, 2013; Noguchi and Stewart 2018; Gluth et al. 2018). However, in most of these computational theories attention plays only a secondary role in explaining context-dependent effect such as the decoy. Thus, we hypothesize that attention could play a more important role in explaining context-dependent preference reversals than previously believed. This thesis aims to seek evidence to support the assumption of a key role of attention in preference reversals as an inner mechanisms of the decision process. In particular, we are going to examine the role of eye-movements, i.e. visual attention, that have been shown to be good proxy for overt attention (Mohler and Wurtz, 1976; Kustov and Robinson 1996).

In the first two studies described in the next chapters, we combined behavioral choice-tasks, eye-tracking and computational modelling to examine the role of visual fixations in two context-dependent changes of preferences during risky decision problems, the framing effect and the decoy effect. Before going into the details of the experimental work which is summarised in the next chapter, I will first review some neurobiological basis of visual-attention and its link with subjective values in value-based decisions, which will provide a background to better understand our hypotheses on the role of visual-attention in context-dependent effects tested in the first two studies of interest. Then, I will discuss the computational modelling approach that will be used to test some of the hypotheses and disclose possible inner mechanism in the decision process. In this computational modelling section, I review the main features of sequential sampling models and the main reasons for adopting such models in decision-making and in particular in our research. Second, I will discuss when this models are not appropriate for the data and what is the best practice to follow when attempting to computationally model empirical data sets.

1.1 Attention

Deciding between options requires a flexible selection of information gathering. In the visual domain, this function is implemented by attention. Visual attention serves to prioritize stimuli processing according to their physical salience ("bottom-up" or exogenous attention) or their relevance to current decision goals on specific rules or motivation factors ("top-down" or endogenous attention), while it discards irrelevant stimuli. At the neuronal level, it has been shown in several electrophysiological studies that attention have different ways to modulate features of the neurons activity improving the signal to noise ratio. First, attention modulates the firing rate of visual neurons enhancing activity of neuronal populations representing the attended stimuli. This modulation has been shown to be multiplicative and the scaling of the neuronal response to be dependent on the similarity between the neuron's preferred stimulus and the attended feature. Second, attention decreases the variability of responses across trials and the correlated variability among neurons. Third, more recent studies have shown that attention can also enhance local gamma frequency (30–60 Hz) synchronization among neurons that encode the attended stimuli (for reviews see Paneri and Gregoriou 2017; Sapountzis and Gregoriou 2018).

Numerous studies have established the prefrontal cortex (PFC) as the candidate brain area in target selection and attentional shifts. For instance, many studies have associated enhancements in frontal eye field (FEF) activity with attention, and FEF appears to have a general role in highlighting locations of behavioral relevance both in exogenous and endogenous attention (Thompson et al., 1997, 2005; Armstrong et al., 2009; Gregoriou et al., 2012). Besides the modulation of neural firing activity, FEF has also shown local changes in gamma frequency synchronization related to spatial attention (Gregoriou et al., 2009b). Although a causal role of the different cortical areas in the different aspects of attention are a matter of current debate, it is well-established a prominent role of PFC neurons in encoding current goals and rules, and facilitate selective processing of information and planning through response modulation in posterior visual areas.

Another well-established role of PFC is its implication in value-based decision making. In particular, activity in the frontal lobe has been shown to reflect the subjective-value of the available options (Kable and Glimcher 2007; Padoa-Schioppa and Assad 2006; Bartra et al. 2013; Rangel and Clithero 2013) and to predict choices during the decision process (Tusche et al. 2010). Also, more recent work has found that patients with damage to this region and adjacent orbitofrontal cortex are more inconsistent when making preference-based choices and violate transitivity suggesting a casual link of PFC and value-based decision processes (Fellow and Farah 2007; Henri-Bhargava et al. 2012; Camille et al. 2011). Thus, it seems clear that attention and subjective values could be closely related in the brain. Also at the behavioral level, recent studies have shown that visual fixations influence value-based choices. In particular, it has been observed that decision-makers choose options they have looked at longer more often than would be predicted by their a priori value ratings of those options alone (Krajbich et al.2010; Krajbich and Rangel 2011, Krajbich et al. 2012) , and that this bias is present even when the duration of fixations is experimentally manipulated (Shimojo et al. 2013; Armel et al. 2008; Tavares et al. 2017). There are not many models in literature that account for this effect of attention on choice frequency in value-based decisions. Notable ones are the attentional drift diffusion model (aDDMs; Krajbich and Rangel 2011) and the mutual inhibition with value-based attentional capture (MIVAC) model. These and other models that make assumption on the allocation of attention during the decision process are reviewed in the supplementary material of study 2 in the appendix of the dissertation. Here, I will mainly focus on sequential sampling models which have become the dominant theory in the cognitive sciences and decision neuroscience.

1.2 Computational Modelling

Like signal detection theory (SDT) (Green and Swets, 1966), the theory of sequential sampling mechanisms starts from the assumption that simple perceptual and cognitive decisions are statistical in nature. This premise comes from the observation that sensory and cognitive systems are inherently noisy and that the decision-making process can be sequential even when all of the information is immediately available. Sequential sampling models (SSMs) have become extremely popular in cognitive sciences in the domain of perceptual decision making for mainly two reasons. First, they can reproduce the observed trade-off speed accuracy of choices, i.e. faster choices are less accurate than slow choices, and they make testable predictions on other aspects of the decision process such as difficulty of task and response time. Second, the core mechanisms of these models, such as evidence accumulation and the decisions threshold crossing, resemble features of neural firing rate activity. More recently, however, both cognitive scientists and decision neuroscientists have become increasingly interested in applications of SSMs to value-based decisions. These decisions, also known as preferential choices, involve affective rather than inferential evaluations and the tasks adopted to to investigate this type of decisions usually employ more complex choice options – i.e. options that are characterized by multiple attributes such as the amount and probability of gambles, or the amount and delay of intertemporal choice options, or the price and quality of consumer products. Some notable examples of SSMs are the drift diffusion model (DDM; Ratcliff, 1978; Ratcliff and McKoon, 2008), the decision field theory (DFT; Busemeyer and Townsend, 1993) and the leaky competing accumulator model (LCA; Usher and McClelland, 2004). An important feature that all of these models have in common is the accumulation to threshold which assumes that the decision mechanism samples information (or accumulates evidence) until a threshold is reached at which point a choice is obtained. In practical terms, a SSM can be formalized as stochastic process that represents the accumulated evidence available to the decision mechanism at a given time. This process will be denoted X(t) in study 1 and 3 . Formally, X(t) is a random variable defined on the probability space of all possible sequences of accumulated information at time t.

The mentioned models, however, can differ in including different features to their models. For instance, they can have one versus two accumulators in the decision rule which can in turn be relative or absolute, or they can differ in the drift rate, i.e. constant or time varying, and also in the presence or absence of inhibition or decay. All of the different features in SSMs make very distinct and testable predictions that can be confirmed or rejected by analyzing experimental data sets. Most of the mentioned features like boundary separations and time-varying drift rates have been widely explored in cognitive sciences (Ratcliff and Smith 2004; Turner et al. 2017; Ratcliff et al. 2018; Voss et al. 2019). However, their need in terms of improving data fitting or their neurobiological validity is still not clear and matter of current debate. Yet, an open question in decision-making research is how to choose the most appropriate model to fit specific data of interest and whether adding extra parameters to the model is a necessary or redundant exercise. We also encountered similar issues when planning and implementing the computational modelling in our experimental studies on visual-attention in context-dependent preference reversals. Thus, we took a less traditional approach on the modelling phase of our research. We proposed that before fitting new experimental data sets few important steps have to be followed. First, testing all the possible predictions of the model to the experimental data set and vice-versa, i.e. testing patterns in the behavioral data and confirming that the model can account for them, increase the probability of the model of being the generating process behind the behavior. Second, performing parameter recovery fitting analysis on the specific parameters resulting from the model fitting assures the researcher that the model and the used fitting method can successfully retrieve the generating parameter of the decision process, and it can distinguish between different magnitudes of these parameters (for a detailed explanation see appendix: study 3). The two steps should be done in the mentioned order and failing in the first one should discourage the researcher from fitting the model to the data. Concerning the first step, this procedure was successful in one out of two studies presented in this dissertation.

In the first study, we investigate the role of visual-attention in the framing effect and we test the hypothesis that visual-attention in our task affects choice frequency according to the attentional drift diffusion model theory (aDDM; Krajbich and Rangel 2011; and see appendix: study 1 for a detailed formalization of the model). The reasons for adopting such model are several. 1) The model does not make any specific assumptions on the allocation of attention which is given as an exogenous variable to the model. 2) The model assumes a multiplicative effect of attention on options' values and a similar effect of multiplicative modulation has been observed in the brain at the neuronal response level (Paneri and Gregoriou 2017; Sapountzis and Gregoriou 2018). 3) The aDDM appeared to be relative simple to estimate and fit to the data. In the third study presented in this dissertation, we show how the aDDM can be approximated by a standard DDM, and thus be fitted and estimated to the data with already existing hierarchical bayesian methods (see appendix: study 3 for a general discussion about the benefits of using hierarchical bayesian methods). As previously discussed however, these reasons are not enough to conclude that the model can be a good approximation of the generating decision process. Thus, we also test in our specific data set that first the predictions of the model are met in the behavioral and eye-gaze empirical data, and second that the model can predict the patterns observed in the data before fitting the model and make any conclusions on the generating process. This approach was successful in our study 1, but the data set in study 2 failed some of the predictions of the aDDM and other models which incorporate attention. The result in study 1, where all the predictions of the aDDM are met by the data set, is certainly not a proof of the fact that the model is the exact generating process mechanism of the data, however it ensures that the model could be a good approximation of the inner decision process. On the contrary, in study 2, where we investigate the role of visual-attention in the decoy effect, we test several different models incorporating visual-attention (including the aDDM) and find that all of them fail in at least one prediction concerning either attention allocation or the interaction choice frequency and visual-fixation patterns. Unfortunately, our data set was not designed to propose a different new model for the decoy effect. Thus, study 2 opens more research questions about the inner computational mechanisms on the interaction between visual-attention and choice frequency in the decoy effect than it can truly answer. Further data and research are needed to answer the following research questions: how does the decoy affect visual-attention allocation in this type of context-dependent effect? And how does visual-attention in turns affect choices? With this example, we show that even though models can in practice fit the experimental data, it is not necessarily informative of the inner mechanisms of the decision process if the predictions of the model do not fully match the patterns in the experimental data. This statement might appear trivial to researcher who are experts in modelling data, however, this does not seem to be often the followed procedure in the field of psychology, cognitive neuroscience and neuroeconomics. Thus, we use our studies also as an example of a possible approach from data to computational modelling that can fail and can be informative of the generating decision process. In the next chapter, I will summarize the three studies that are the core work behind this dissertation and that can be found in the appendix section of the manuscript.

Chapter 2: Overview of the studies

As I describe in the previous chapter, this thesis primarily aims to seek evidence for a fundamental role of attention during the decision-making process as a common inner mechanism among different context-dependent preference reversal effects. In order to rule out how attention, in particular visual-attention, affects choice frequency and response time distributions, and how attention is allocated differently depending on the different context, we measured eye-movements in two lottery task contexts and use computational modelling tools to further investigate the inner processes. In study 1 summarized above, we investigate the framing effect in a lottery choice-task and we examine the mechanisms through which visual-attention can influence choices and generate the recorded behavioral framing effect in an attentional drift diffusion framework. After testing and confirming the predictions of the aDDM in our data set, we directly fit the model to choice and response time distributions with a hierachical bayesian version of the aDDM (HaDDM).

The method we implemented to perform parameter recovery fitting analysis and fitting the model to the experimental data is described in study 3.

Study 2 investigate the role of attention for the decoy effect in a similar lottery task to study 1. In this study eye-movements are also recorded, and different models which make predictions about visual-attention allocation and its interaction with choice frequency are tested.

2.1 Study 1: A mechanistic foundation of the role of visual-attention in the framing effect

Background

Subjective preferences have been shown to change across different contexts even when the options are kept constant. For example, Kahneman and Tversky have shown that humans tend to prefer sure options over risky prospects, when these options are presented as gains, but this aversion to risk is diminished if the sure options are presented as losses. This effect of framing on choices occurs even when both decision problems are identical in terms of possible outcomes (Kahneman Tversky, A., 1979; Tversky and Kahneman, 1981). The framing effect is a well documented example of preference reversal in which individuals seem to change their subjective valuation of the options in the choice set. However, the neurological and internal mechanisms that lead people to change their choices depending on the different frames are still unclear. Why should mere changes in the description of an objectively identical relationship between choices and outcomes affect decision-making?

Here, we examined the hypothesis that changes in framing cause changes in the allocation of visual-attention to the different options, and in turns visual-attentional changes give rise to changes in the decision process. To build mechanistic and computational supports for this thesis, we adopted an attentional drift diffusion framework (i.e. an attentional drift diffusion model or aDDM). According to this model, visually guided attention temporarily biases the decision process in favour of the alternative that is attended (Ashby et al., 2016; Cavanagh et al., 2014; Konovalov and Krajbich, 2016; Kovach et al., 2014; Krajbich and Rangel, 2011; Krajbich et al., 2010; Lim et al., 2011; Schonberg et al., 2014; Shimojo et al., 2003; Stewart et al., 2016; Towal et al., 2013; Vaidya and Fellows, 2015). The advantage of using a particular framework such as the aDDM can be found in the fact that it provides accurate and quantitative predictions that can be tested in our experimental data. In particular, the aDDM makes specific predictions about the relationship between options' values, visual fixations, reaction time and choices. The key aim of this study was to adopt such quantitative predictions to make inferences about how and whether changes in preferences are influenced by visual-guided attention in the context of the framing effect. We documented that in decision making under risk the framing of sure alternatives as a gain – as opposed to a loss – induces a visual-attentional advantage for the sure option relative to the risky one which is accompanied with an increase in the choice probability of the sure option, i.e., an increase in risk aversion. Based on the aDDM framework, we proposed an explanation for the framing effect that is exclusively dependent on a reallocation of visual-attention through changes in the parameters of the model that do not involve a change in the valuation of the options in the choice set. We tested our hypotheses with a hierarchical Bayesian modelling approach (HaDDM), and find that frames have an impact primarily on the initial bias of the evidence accumulation process that, combined with changes in the allocation of visual-attention, can fully explain the observed changes in risk aversion. Our results suggest that the main drivers of framing effects in decision-making are shifts in attention allocation that alter the inner mechanisms of the evidence accumulation process.

Methods

We asked 30 participants to perform a series of binary decisions between two options, a gamble and sure option, in two differently framed conditions while we simultaneously recorded their gaze patterns with eye-tracker. In one condition the sure alternative on the decision screen was framed as a gain and in the other condition the same sure option was framed as a loss as in De Martino et al. (2006). Each trial started with a screen that showed the subject's initial monetary endowment for that trial. Then, a decision screen that showed the gamble and the sure option appeared. Each option was presented to the subjects as a combination of a rectangle and a pie-chart (as an example see appendix study 1 figure A.1) that the subjects trained with two extensive training phases before the choice task in order to recongnize and learn the meanings of the shapes and colors. After the binary choice task, the subjects had to perform a elicitation task in which they were shown all the gambles that they encountered during the choice task. In each trial a gamble and a list of sure amounts of money was shown to the participant (as an example see appendix study 1 figure A.2). The subject had to make 10 decisions in each trial between the gamble displayed on the left hand-side of the screen and 10 sure amounts of money. The sure amounts of money were displayed in a decreasing order from a maximum value smaller but close to the amount of money that they could win if they choose the gamble to a minimum value close to zero. For each gamble, which was constant across the two frame conditions, i.e. gain and loss, we used the elicitation task to estimate its subjective value or certainty equivalence (CE). The subjects were incentivized for the experiment with real money. They were informed that one of the two tasks – i.e. the choice task or the elicitation task – was randomly selected and implemented as a payment at the end of the experiment. Concerning the modelling, we fitted the HaDDM to the data with the method described in study 3 in the appendix. In the hierachical bayesian method, we make the standard assumption that individuals are members of a normally distributed population and assign a normal prior for each individual level parameter. Furthermore, before fitting the model we ran recovery fitting analysis to make sure the model was able to correctly fit the data.

Results and conclusions

At the behavioral level, we replicated the framing effect on choices, i.e. participants on average chose the sure option in the gain condition more often than in the loss condition. Then, we tested the hypotheses of the aDDM on the relationship between eye-gaze fixations, options' values, choices and reaction time on our experimental data and find that the predictions of the model were confirmed. First, as expected subjects looked more often at the sure option in the gain condition compared to the loss condition revealing a framing effect at the fixation level which was also significantly and highly correlated with the framing effect on choices at the individual level. Secondly, visual-attention seems to impact choice frequency independently of the framing condition when analysing the impact on choices of fixation biases towards an option in both gain and loss conditions. Thus, both at the individual and at the trial level visual-attention is significantly correlated with choice frequency and the framing effect. Also, we found that the predictions of the model about the relation between reaction time and fixation biases are in line with the model. We registered faster decisions for the sure option in the gain condition compared to the loss condition, and faster response time in choosing the sure option compared to the gamble at higher expected value trials. The previous evidence provides support for the view that the framing effect is an attention-driven phenomenon. It appears that the gain frame directs more attention towards the sure option which then enhances the probability of choosing the sure option more frequently and more quickly. With the evidence collected thus far however, we could not claim that the link between choice frequency and visual-attention goes from fixations to choices, or vice-versa, but separately analysing fixation time and eye-gaze movements we found some additional evidence supporting the hypothesis that the decision makers in our experiment are more likely to choose what they look at and not vice-versa. In fact, we found that both middle fixation duration and first fixation duration do not correlate with the options' values, and also we do not find a significant dependency between visual fixation time and value differences in our data. This evidence seems to suggest a prominent role of visual-attention on choices that is driven by the condition. According to the aDDM framework, the frame changes where people look, and this change in visual gazes can account for part of the changes in choices across conditions. We then were interested in investigating how framing also affects the different aspects of the evidence accumulation process, i.e. the parameters of the aDDM. In particular, we noticed that changes in visual fixation patterns alone are unlikely to be the only source through which the framing effect is generated, since there is still some unexplained variance in our data that cannot be fully explained by visual-attention alone. Thus, we expect other parameters of the aDDM to vary across conditions. We fitted the HaDDM to our data and estimated the parameters of the model. The hierarchical estimation of the mean population parameters at the group level allowed us to make inference about the parameters accounting for variation coming from sources like individual differences. Thus, we could benefit from this method to investigate the group level differences in parameters across gain and loss fitting the model separately for the two conditions. We found that the main difference in parameters between loss and gain was in the initial bias and in the drift rate parameters. Additionally, the attentional discount parameter was slightly lower – i.e. higher discount for the unattended option in the gain condition than in the loss condition and, in addition, with lower variability – i.e narrower distribution in the gain condition. We also split subjects in two groups based on whether they showed a framing effect below or above the median and analysed the parameters for the two groups at the individual level. For the above median framing effect group of subjects the initial bias, the drift rate constant and the discount factor were significantly different across conditions, whereas they were not for the below median framing effect group of individuals. Ultimately, we ran a logistic mixed-effect regression for subject specific slope and constant of the probability of choosing the sure option as a function of condition, proportion of fixation time to the sure option, value difference, expected value, and, in addition, for every subjects their mean value for each parameter of the aDDM for gain, one value for loss condition and all the interactions with proportion of fixation time to the sure option. Notably, the results show that the frame condition did not have significant effect on the probability of choosing the sure option that was then completely accounted by the visual gazes, the changes in the parameters and their interactions. To conclude, we found that gaze patterns, changes in drift rate and initial bias, and their interactions can completely account for the framing effect without changing the valuation of the options. Thus, our results suggest a prominent role of visual-fixation time in the framing effect that interacts with the decision process changing the speed of the accumulation and the initial advantage of the sure option across frames.

2.2 Study 2: Is the decoy effect an attention-driven phenomenon?

Background

The decoy effect is a well documented example of a preference reversal in which individuals seem to change their subjective valuation of two options (A vs B) when a third, irrelevant, alternative (D) is introduced in the choice-set. Although the decoy effect has been known in Economics since 40 years and also more recently has been widely investigated in psychology and in neuroscience, which tried to unveil the brain and psychological processes during this type of decisions, there is still scarce consensus on the driving mechanisms for such effect and the conditions under which decision makers exhibit different forms of this violation are a matter of current debate. There have been many computational modelling attempts to explain the various types of decoy effects (Kahneman and Tversky, 1984; Busemeyer and Townsend, 1993; Hotaling et al., 2010; Roe et al., 2001; Usher and McClelland, 2004; Bhatia, 2013; Noguchi and Stewart 2018; Gluth et al. 2018), and more recently the idea that attention could play a role in the decision process have been incorporated in some of models (Krajbich and Rangel, 2011; Tsesos, Chater and Usher 2012; Gluth, Spektor and Rieskamp 2018; Nouguchi and Stewart 2014). However, in most of these computational theories attention plays a secondary role in explaining context-dependent effect such the decoy. Additionally, in very few studies that investigate the decoy effect (Nouguchi and Stewart 2014; Gluth et al. 2018) attention has been fully recorded. In the present work, we examined the hypothesis that attention plays a more important role in explaining the decoy effect than previously believed. The study aimed to seek evidence to support the assumption that introducing a third alternative in the choice-set causes changes in the allocation of visual-attention to the options, and in turns visual-attention changes give rise to changes in the decision process. Thus, we combined a behavioral paradigm, eye-tracking and computational techniques to examine the role of visual fixations as a proxy for attention in decoy-dependent changes of preferences during risky decision problems. We developed a lottery task in which subjects performed a series of trinary decisions between risky options while their gaze movements were recorded. One of the options was always the irrelevant decoy, i.e. always dominated in money and probability by one of the other two options which we refer here as the target. We investigated and replicated the attraction effect. Although the decoy was essentially never chosen it had a considerable effect on the relative allocation of attention to the two relevant options. If the decoy was similar to option A, i.e. option A is the target, it induced frequent comparisons between the decoy and this option and thus strongly increased the relative attention to option A. Additionally, we recorded an effect on choice frequency of looking time to the decoy. Surprisingly, the probability of choosing the target was significantly influenced by the time participants spent attending the decoy. We investigated this result further and find evidence suggesting that that the decoy alters attention allocation to the choice set making the target option the center of attention. From our result, it seems reasonably clear that visual-attention plays a major role in the decoy effect. To further investigate the mechanism with which the decoy influences visual-attention allocation and visual-attention influences choice frequency, we tested the predictions of models that assumes a role of attention in the decoy effect. We found that only very few features in the examined models seem to be in line with our eye-gaze data, suggesting a more complicated and still unsolved role of attention in the decoy effect. Our results suggest a fundamental role of the decoy in attracting and shaping attention allocation which was not previously recorded. The decoy causally manipulates the attention allocated by the decision maker on options in the choice set and this shifts in attention are strictly related to choice frequency in very specific ways that have not been previously accounted for.

Methods

40 healthy subjects performed a series of trinary decisions between three lotteries, option A, option B and a decoy, while we simultaneously recorded their gaze patterns with an eye-tracker. Each option, if chosen, would always give an amount of money with a certain probability p or nothing with probability 1-p. During the all task, option A was the option with high money low probability trade off and B was the low money high probability option, whereas the decoy could be either of type A, i.e. low money high probability, or of type B, i.e. high money low probability. Irrespectively of its type, the decoy was always a dominated option either by option A or by option B and the decoy determines the condition type. Namely, in half of the trials the decoy was always smaller in both attributes, i.e. money and probability, than option A – i.e. condition 1, see Figure B1a in appendix for an

example –, and in the other half of the trials it was smaller in both attributes than option B – i.e. condition 2 in figure B1a in appendix. Whenever the decoy is of type A we refer to option A as the target option and option B as the competitor. Vice-versa, when the decoy is of type B, option B is the target option and A the competitor. During the presentation of the decision screen, subjects were not shown numbers and words; they could only see shapes and colors of which were extensively trained before the choice task to learn and understand the meaning. Our participants were incentivized for the experiment with real money. They were informed that they would receive the average amount of money of all the trials realized choices that they made. Furthermore, subjects were instructed to treat each decision separately and independently of the others.

Results and conclusions

At the behavioral level, we observed the predicted attraction effect, i.e. the target was chosen significantly more often than the competitor. Then, we tested the hypothesis that also at the fixations level we would expect to see a decoy effect, i.e. a bias in attention towards the target option. Thus, we asked whether the decoy not only changed the choice frequency between A and B, but whether it also changed the allocation of attention between the two options. As expected, there was a significant bias in fixation time for option A or B depending on whether the decoy is of type A or B respectively, thus resulting in a decoy effect on fixations. In addition to the strong effect of a fixation bias on choice frequency, we also noticed that the decoy effect can be reversed for trials in which the fixation bias is in turn reversed, namely trials in which the competitor was attended longer than the target had also a higher choice probability for the competitor than the target. We then investigated why and how the the decoy makes the target to be attended longer, and find that the decoy forced more eye-gaze transitions from and to the target. The looking time of the decoy in a trial decreased in time, suggesting that once the dominance relation between target-decoy had been resolved the decision maker did not need to look at the decoy. Thus, we speculated that the decoy makes the target the center of attention attracting more evidence in favour of choosing the target. Another effect that surprisingly we recorded is that the decoy effect on choices, i.e. the difference in probability of choosing the target and the competitor, in our data was significantly increased by the proportion of fixation time to the decoy. Thus, only looking at the decoy increased the probability of the target to be chosen. Also, these two main effect of proportion of fixation time to the target and to the decoy were highly correlated at the individual level to the decoy effect on choices. In a further analysis, we investigated the predictions on visual-attention allocation, i.e. the probability of attending an option, and the predictions on the relation between visual-attention and choice frequency of some of the models in literature that incorporate attention in their framework. This analysis allowed us to test which are the most likely mechanisms through which visual-attention operates to influence choices in the decoy effect. It turned out that most the predictions of the models could not be found or were only partially met in our experimental data. Overall our results suggest very strong effect of visual-attention on choices that, however, seems to be more complicated that previously modelled. Also, it appeared from our results that the decoy its-self and not necessarily the values or the distance between options was the main driver of attention allocation. Namely, the different types of decoy made the decision maker to allocate attention differently during the decision process, and this attention allocation bias seemed to have a strong impact on choices. Unfortunately, our experiment was not designed to rule out the dependent variables that impact visual-attention allocation to options, thus further analysis and different designs are needed to better investigate the probability of attending the option in the choice-set and better understand the visual-attention influence on choice frequency in the decoy effect.

2.3 Study 3: Simple method to fit hierarchical Bayesian piecewise time-constant drift diffusion models.

Background

Sequential sampling and accumulation to threshold principles have proven to be important and highly influential in modelling value-based and perceptual decision making in the social and biological sciences. To date, most examples of this class of models that have been fit to empirical data have assumed a constant evidence accumulation or drift rate throughout the choice process – e.g. the standard drift diffusion model (DDM; Ratcliff, 1978) or the leaky competing accumulators model (LCA; Usher and McClelland, 2004). The reasons for maintaining this assumption has been the practical considerations relating to the computational complexity and time required to fit models that relax this assumption and the rise of hierarchical Bayesian estimation methodologies (Vandekerckhove, Tuerlinckx, and Lee, 2011; Wiecki, Sofer, and Frank, 2013; Wabersich and Vandekerckhove, 2014) which often require an analytical solution of the equations within the model. Despite the clear advantages of time-varying drift rate sequential sampling processes in modelling different type of decision strategies and inner mechanisms, it often remains theoretically and computationally challenging to estimate and implement such formulations. However, most of the time-changing drift rate diffusion processes found in literature can be capture with a drift rate that is time-constant in fixed or alternating intervals, i.e. it is piecewise time-constant. Diffusion models with piecewise time-constant drift rates (pcDDM) can be approximated with a standard drift diffusion model (DDM) for which the drift rate is constant in time.

Methods

In this study, we illustrated how to transform a pcDDM into a standard DDM with a fully time-constant drift rate allowing for the use of existing hierarchical Bayesian estimation tools that were developed for the DDM. We showed that a piecewise time-constant drift rate can be expressed as a constant drift that is a function of the first passage time - i.e reaction time - and the times for each constant interval. To demonstrate the efficacy and the benefit of this method, we applied it to two practical examples, the attentional drift diffusion model (aDDM; Krajbich and Rangel, 2011; Krajbich et al., 2010) and the time-varying sequential sampling (tSSM; in preparation). We derived the mathematical formulations of the method for these two examples, and we applied them to simulated and real data sets. Our results show that the method allows us to quickly and accurately recover parameters from simulations and fit empirical data sets in a hierarchical Bayesian fashion. In the study, we derived mathematically the drift rate of a general pcDDM and then applied it to two example models, the aDDM and the stDDM.

Then, we used two experimental data sets to test our models. The first data set was taken from the gain condition of the lottery task described in study 1, in which we used an eye-tracker to record gaze-movement data that are given as an input to the aDDM.

Concerning the second data set, eighty-six healthy subjects participated in a food task choice experiment. The experimental design was divided into two parts: a rating phase and a decision phase. During the rating phase participants had to rate 180 food items for healthiness and tastiness. After the ratings, subjects had to perform a series of binary food-choice decisions in which two food items were presented on the computer screen and they had to choose the food that they would like to receive at the end of the experiment. In addition, subjects had to express their willingness to try to eat, i.e. choose, healthy. Participants who were not willing to try to eat healthy were subsequently excluded from the analysis.

Concerning the recovery and the data fitting analysis, we implemented the aDDM and the stDDM models with the derived method in R and Jags. We made use of the existing function in Jags to fit a standard DDM, dwieners(). The details on the code for simulations, recovery fitting and data fitting analyses can be found in the appendix and in the github repository at https://github.com/galombardi/method_HtSSM_aDDM.

Results and conclusions

We showed that any pcDDM model, i.e. any sequential sampling model with a piecewise time-constant drift rate, can be written as a DDM with constant drift rate that depends on the reaction time and on the time intervals where the drift rate of the pcDDM is constant.

Then, we tested the recovery fitting ability of two different pcDDMs, the aDDM and the stDDM. First, we ran a recovery fitting analysis where all subjects had the same parameters values. We observed that at the group and at the individual level, the hierarchical aDDM and the hierarchical stDDM were both able to recovery different values of the parameters in the drift rate, i.e. the attentional parameter θ and the time parameter s respectively, without affecting the other parameters of the DDM. This also hold when performing a recovery fitting analysis with heterogeneous parameters where each subject had parameters for the simulations that were randomly drawn from gaussian distributions. Second, both models performed well when we ran a recovery fitting analysis using the parameters coming from the experimental data fitting. Namely, we first fitted the two data sets described above with the hierarchical aDDM and the hierarchical stDDM, then we used for each subject the mean parameters given by the fitting to simulate the models and try to recover the original fitted parameters. We also ran an extra analysis where we used the same fitted parameters for the DDM but try to recover different values of the drift parameters such as the attentional discount factor θ in the aDDM and the time parameter s in the stDDM. Ultimately, we compared the reaction time and choice distributions of the real data with the fitted ones and confirmed that the models could correctly reproduce the experimental data.

Our results confirm that both mathematically and computationally every pcDDMs can be expressed as a standard DDM given the reaction time and the time-constant intervals. We also showed the importance of recovery fitting analysis in particular when dealing with new data sets, and we suggested a before-fitting confirmatory parameter recovery procedure to test whether a particular data set can be fitted with the model of interest. The first step in such a recovery test is to check the ability of the model and fitting procedure to recover known generating parameters. This can be done by simulating the data with the parameters obtained when originally fitting the empirical data, and then re-fitting the model to the simulated choice and response times. The second step of the recovery test consists in making sure that the model and the fitting procedures are able to accurately distinguish between different generating parameter values. Specifically all the simulating parameters but one have to be taken from the fitting of the empirical data as in the first step. The one remaining parameter is then varied across a range of plausible values to test. This procedure ensures that the model can recover different magnitudes of that specific parameter of interest under the real data conditions.

Chapter 3: Discussion

One goal of decision making is to understand the inner and generating processes that give rise to preference reversals in context-dependent value-based decisions. In this thesis, we explore the hypothesis that a key factor and common mechanism that can drive different context-dependent choices is attention. Growing literature has identified a prominent role of attention in perceptual and value-based decision-making (Krajbich et al. 2010; Krajbich and Rangel 2011; Krajbich and Rangel 2012; Towal et al. 2013), however, attention has never been recorded and investigated in different context-dependent preference reversal effects.

Here, I explored experimentally and computationally how visual-attention, which has been shown to be a good proxy for overt attention (Mohler and Wurtz, 1976; Kustov and Robinson 1996), is allocated and it affects choice frequency in two popular context-dependent effects, i.e. the framing effect and the decoy effect.

Study 1 investigated the framing effect in an attentional drift diffusion framework, providing evidence for an explanation of this preference reversal only through changes in visual-attention allocation and its interactions with the sequential sampling decision process. Without a change in the valuation of the options across frames, we could provide evidence for a possible decision process mechanism where the context influences attention and attention has in turns an impact on choice frequency. Contrarily to an other stream of research in the context-dependent decision literature which identifies preference reversals as the result of an encoding representation of values in a normalized form (Louie et al. 2013; Landry and Webb, 2019), we did not find evidence for such normalization effects in our studies. Absolute options' values appeared to have an impact on choice frequency which cannot be accounted for by such adaptation effects. Thus, we took a drift diffusion modelling approach and we developed a simple, yet efficient, method to perform the fitting procedure in a hierarchical Bayesian fashion.

Study 3 showed that an aDDM model can be easily formalized as a standard DDM thus improving and simplifying the computational workload. In this study, we also emphasized the importance of carefully perform parameter recovery fitting analyses before applying the model or the method to empirical data. Also, our approach in study 1 showed an example of how to rigorously test all the quantitative predictions and assumptions of a model before fitting the model to the data.

As a further example, study 2 showed instead a failure of this testing procedure and led us to the decision of not fitting any existing models to our experimental lottery data. Specifically, study 2 investigated the role of visual-attention in a decoy effect lottery choice-task, i.e. the attraction effect. This study underlined the need for a better mechanistic understanding on the role of visual-attention in the decoy effect. As for the framing effect, we found a prominent role of visual-attention in influencing choice frequency in the attraction effect, however, none of the exiting models seemed to be fully in line with our experimental data set. Most of the theories we tested, such as multialternative decision field theory (MDFT; Busemeyer and Townsend, 1993; Hotaling et al., 2010; Roe et al., 2001), the multiattribute leaky competing accumulator (MLCA; Usher and McClelland, 2004), the associations and accumulation model (AAM; Bhatia, 2013), the multialternative decision by sampling (MDbS, Noguchi and Stewart 2018), or the mutual inhibition value-based attention capture (MIVAC, Gluth et al. 2018), make precise predictions about how attention should be allocated among options in the choice set.

Some models such as the AAM, the MDbS and MIVAC assumed that the probability of attending an option should depend on the overall value of the attributes, the value distance from other options or on the option's integrated value, respectively. We found that only partially this was confirmed in our data. More specifically, the degree and the sign of the dependency on the mentioned features of the choice set varied with the option type, i.e. either the decoy, the target or the competitor, which seemed to be the main factor driving attention as if the decision-maker would recognize the type of option and change attention allocation accordingly.

Furthermore, in study 2 we observed different effects of attention never recorded before. We first noticed that the eye-gaze transition probability strictly depended on the option's type, i.e. either the decoy, the target or the competitor. Secondly, we found an effect of looking time to the decoy onto the probability of choosing the target which was a priori unexpected.

To summarize, this work shows how mechanistically and behaviourally visual-attention could have a fundamental role in explaining context-dependent preference reversal effects without directly changing the valuation of the options. This work also aim to present computational modelling case studies and show a possible efficient procedure to follow before fitting and applying models (or new methods) to the fitting of empirical data sets.

Bibliography

Anderson, B. A., P. A. Laurent, and S. Yantis (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences 108*(25), 10367–10371.

Armel, K. C., A. Beaumel, and A. Rangel (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision making 3*(5), 396–403.

Armstrong, T. and B. Olatunji (2009). What they see is what you get: Eye tracking of attention in the anxiety disorders. *Psychological Science Agenda 23*(3).

Ashby, N. J. and T. Rakow (2016). Eyes on the prize? evidence of diminishing attention to experienced and foregone outcomes in repeated experiential choice. *Journal of Behavioral Decision Making 29*(2-3), 183–193.

Awh, E., E. K. Vogel, and S.-H. Oh (2006). Interactions between attention and working memory. *Neuroscience 139*(1), 201–208.

Bartra, O., J. T. McGuire, and J. W. Kable (2013). The valuation system: a coordinate-based meta-analysis of bold fmri experiments examining neural correlates of subjective value. *Neuroimage 76*, 412–427.

Berkowitsch, N. A., B. Scheibehenne, and J. Rieskamp (2014). Rigorously testing multialternative decision field theory against random utility models. *Journal of Experimental Psychology: General 143*(3), 1331.

Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological review 120*(3), 522.

Bhatia, S. (2014). Sequential sampling and paradoxes of risky choice. *Psychonomic Bulletin & Review 21*(5), 1095–1111.

Bird, G. D., J. Lauwereyns, and M. T. Crawford (2012). The role of eye movements in decision making and the prospect of exposure effects. *Vision Research 60*, 16–21.

Bitzer, S., H. Park, F. Blankenburg, and S. J. Kiebel (2014). Perceptual decision making: drift-diffusion model is equivalent to a bayesian model. *Frontiers in human neuroscience 8*, 102.

Bogacz, R., M. Usher, J. Zhang, and J. L. McClelland (2007). Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice. *Philosophical Transactions of the Royal Society B: Biological Sciences 362*(1485), 1655–1670.

Brody, C. D. and T. D. Hanks (2016). Neural underpinnings of the evidence accumulator. *Current opinion in neurobiology 37*, 149–157.

Bronfman, Z. Z., N. Brezis, and M. Usher (2016). Non-monotonic temporal-weighting indicates a dynamically modulated evidence-integration mechanism. *PLoS computational biology 12*(2).

Brown, S. D. and A. Heathcote (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive psychology 57*(3), 153–178.

Busemeyer, J. R., S. Gluth, J. Rieskamp, and B. M. Turner (2019). Cognitive and neural bases of multi-attribute, multi-alternative, value-based decisions. *Trends in cognitive sciences 23*(3), 251–263.

Busemeyer, J. R. and J. T. Townsend (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review 100*(3), 432.

Callaway, F. and T. Griffiths (2019). Attention in value-based choice as optimal sequential sampling.

Camille, N., C. A. Griffiths, K. Vo, L. K. Fellows, and J. W. Kable (2011). Ventromedial frontal lobe damage disrupts value maximization in humans. *Journal of Neuroscience 31*(20), 7527–7532.

Cavanagh, J. F., T. V. Wiecki, A. Kochar, and M. J. Frank (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology: General 143*(4), 1476.

Chau, B. K., N. Kolling, L. T. Hunt, M. E. Walton, and M. F. Rushworth (2014). A neural mechanism underlying failure of optimal choice with multiple alternatives. *Nature neuroscience 17*(3), 463.

Churchland, A. K., R. Kiani, and M. N. Shadlen (2008). Decision-making with multiple alternatives. *Nature neuroscience 11*(6), 693.

Clay, S. N., J. A. Clithero, A. M. Harris, and C. L. Reed (2017). Loss aversion reflects information accumulation, not bias: a drift-diffusion model study. *Frontiers in psychology 8*, 1708.

De Martino, B., D. Kumaran, B. Seymour, and R. J. Dolan (2006). Frames, biases, and rational decision-making in the human brain. *Science 313*(5787), 684–687.

Denwood, M. J. et al. (2016). runjags: An r package providing interface utilities, model templates, parallel computing methods and additional distributions for mcmc models in jags. *Journal of Statistical Software 71*(9), 1–25.

Diederich, A. and P. Oswald (2014). Sequential sampling model for multiattribute choice alternatives with random attention time and processing order. *Frontiers in human neuroscience 8*, 697.

Ditterich, J. (2006). Stochastic models of decisions about motion direction: behavior and physiology. *Neural networks 19*(8), 981–1012.

Donoso, M., A. G. Collins, and E. Koechlin (2014). Foundations of human reasoning in the prefrontal cortex. *Science 344*(6191), 1481–1486.

Fellows, L. K. and M. J. Farah (2007). The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se? *Cerebral cortex 17*(11), 2669–2674.

Forstmann, B. U., R. Ratcliff, and E.-J. Wagenmakers (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual review of psychology 67*, 641–666.

Frederick, S., L. Lee, and E. Baskin (2014). The limits of attraction. *Journal of Marketing Research 51*(4), 487–507.

Glimcher, P. W. and E. Fehr (2013). *Neuroeconomics: Decision making and the brain*. Academic Press.

Gluth, S., J. M. Hotaling, and J. Rieskamp (2017). The attraction effect modulates reward prediction errors and intertemporal choices. *Journal of Neuroscience 37*(2), 371–382.

Gluth, S., M. S. Spektor, and J. Rieskamp (2018). Value-based attentional capture affects multi-alternative decision making. *Elife 7*, e39659.

Gold, J. I. and M. N. Shadlen (2007). The neural basis of decision making. *Annual review of neuroscience 30*.

González-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review 109*(1), 137.

Gottlieb, J., M. Hayhoe, O. Hikosaka, and A. Rangel (2014). Attention, reward, and information seeking. *Journal of Neuroscience 34*(46), 15497–15504.

Green, D. M., J. A. Swets, et al. (1966). *Signal detection theory and psychophysics*, Volume 1. Wiley New York.

Gregoriou, G. G., S. J. Gotts, and R. Desimone (2012). Cell-type-specific synchronization of neural activity in fef with v4 during attention. *Neuron 73*(3), 581–594.

Gregoriou, G. G., S. J. Gotts, H. Zhou, and R. Desimone (2009a). High-frequency, long-range coupling between prefrontal and visual cortex during attention. *science 324*(5931), 1207–1210.

Gregoriou, G. G., S. J. Gotts, H. Zhou, and R. Desimone (2009b). Long-range neural coupling through synchronization with attention. *Progress in brain research 176*, 35–45.

Hanks, T. D., J. Ditterich, and M. N. Shadlen (2006). Microstimulation of macaque area lip affects decision-making in a motion discrimination task. *Nature neuroscience 9*(5), 682–689.

Hanks, T. D. and C. Summerfield (2017). Perceptual decision making in rodents, monkeys, and humans. *Neuron 93*(1), 15–31.

Heath, T. B., S. Chatterjee, and K. R. France (1995). Mental accounting and changes in price: The frame dependence of reference dependence. *Journal of consumer research 22*(1), 90–97.

Henri-Bhargava, A., A. Simioni, and L. K. Fellows (2012). Ventromedial frontal lobe damage disrupts the accuracy, but not the speed, of value-based preference judgments. *Neuropsychologia 50*(7), 1536–1542.

Holmes, W. R. and J. S. Trueblood (2018). Bayesian analysis of the piecewise diffusion decision model. *Behavior research methods 50*(2), 730–743.

Holmes, W. R., J. S. Trueblood, and A. Heathcote (2016). A new framework for modeling decisions about changing information: The piecewise linear ballistic accumulator model. *Cognitive psychology 85*, 1–29.

Hotaling, J. M., J. R. Busemeyer, and J. Li (2010). Theoretical developments in decision field theory: Comment on tsetsos, usher, and chater (2010).

Hu, J. and R. Yu (2014). The neural correlates of the decoy effect in decisions. *Frontiers in behavioral neuroscience 8*, 271.

Huber, J., J. W. Payne, and C. Puto (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research 9*(1), 90–98.

Huber, J., J. W. Payne, and C. P. Puto (2014). Let's be honest about the attraction effect. *Journal of Marketing Research 51*(4), 520–525.

Hübner, R., M. Steinhauser, and C. Lehle (2010). A dual-stage two-phase model of selective attention. *Psychological review 117*(3), 759.

Itti, L. and C. Koch (2001). Computational modelling of visual attention. *Nature reviews neuroscience 2*(3), 194–203.

Johnston, W. A. and V. J. Dark (1986). Selective attention. *Annual review of psychology 37*(1), 43–75.

Jung, K., J. Jeong, and J. D. Kralik (2019). A computational model of attention control in multi-attribute, context-dependent decision making. *Frontiers in computational neuroscience 13*.

Kable, J. W. and P. W. Glimcher (2007). The neural correlates of subjective value during intertemporal choice. *Nature neuroscience 10*(12), 1625–1633.

Kahneman, D. and A. Tversky (1981). The simulation heuristic. Technical report, Stanford Univ CA Dept of Psychology.

Keitel, C., S. K. Andersen, C. Quigley, and M. M. Müller (2013). Independent effects of attentional gain control and competitive interactions on visual stimulus processing. *Cerebral Cortex 23*(4), 940–946.

Khaw, M. W., P. W. Glimcher, and K. Louie (2017). Normalized value coding explains dynamic adaptation in the human valuation process. *Proceedings of the National Academy of Sciences 114*(48), 12696–12701.

Kim, B., D. Seligman, and J. W. Kable (2012). Preference reversals in decision making under risk are accompanied by changes in attention to different attributes. *Frontiers in neuroscience 6*, 109.

Kleiner, M., D. Brainard, and D. Pelli (2007). What is new in psychophysics toolbox. *Perception 36*.

Konovalov, A. and I. Krajbich (2016). Gaze data reveal distinct choice processes underlying model-based and model-free reinforcement learning. *Nature communications 7*(1), 1–11.

Kovach, C. K., M. J. Sutterer, S. N. Rushia, A. Teriakidis, and R. L. Jenison (2014). Two systems drive attention to rewards. *Frontiers in psychology 5*, 46.

Krajbich, I. (2019). Accounting for attention in sequential sampling models of decision making. *Current opinion in psychology 29*, 6–11.

Krajbich, I., C. Armel, and A. Rangel (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature neuroscience 13*(10), 1292.

Krajbich, I., D. Lu, C. Camerer, and A. Rangel (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in psychology 3*, 193.

Krajbich, I. and A. Rangel (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences 108*(33), 13852–13857.

Kustov, A. A. and D. L. Robinson (1996). Shared neural control of attentional shifts and eye movements. *Nature 384*(6604), 74–77.

Landry, P. and R. Webb (2019). Pairwise normalization: A neuroeconomic theory of multi-attribute choice. *Available at SSRN 2963863*.

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical bayesian models. *Journal of Mathematical Psychology 55*(1), 1–7.

Leong, Y. C., B. L. Hughes, Y. Wang, and J. Zaki (2019). Neurocomputational mechanisms underlying motivated seeing. *Nature human behaviour 3*(9), 962–973.

Lim, S.-L., J. P. O'Doherty, and A. Rangel (2011). The decision value computations in the vmpfc and striatum use a relative value code that is guided by visual attention. *Journal of Neuroscience 31*(37), 13214–13223.

Louie, K., M. W. Khaw, and P. W. Glimcher (2013). Normalization is a general neural mechanism for context-dependent decision making. *Proceedings of the National Academy of Sciences 110*(15), 6139–6144.

Luce, R. D. (1959). Response latencies and probabilities. *Mathematical methods in the social sciences*, 298–311.

Milosavljevic, M., J. Malmaud, A. Huth, C. Koch, and A. Rangel (2010). The drift diffusion model can account for the accuracy and reaction time of value-based choices under high and low time pressure. *Judgment and Decision Making 5*(6), 437.

Mohler, C. W. and R. H. Wurtz (1976). Organization of monkey superior colliculus: intermediate layer cells discharging before eye movements. *Journal of neurophysiology 39*(4), 722–744.

Mohr, P. N., H. R. Heekeren, and J. Rieskamp (2017). Attraction effect in risky choice can be explained by subjective distance between choice alternatives. *Scientific reports 7*(1), 1–10.

Moran, J. and R. Desimone (1985). Selective attention gates visual processing in the extrastriate cortex. *Science 229*(4715), 782–784.

Mormann, M. M. and C. Frydman (2016). The role of salience and attention in choice under risk: An experimental investigation. Technical report, Working paper.

Mullett, T. L. and N. Stewart (2016). Implications of visual attention phenomena for models of preferential choice. *Decision 3*(4), 231.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology 47*(1), 90–100.

Noguchi, T. and N. Stewart (2014). In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition 132*(1), 44–56.

Noguchi, T. and N. Stewart (2018). Multialternative decision by sampling: A model of decision making constrained by process data. *Psychological review 125*(4), 512.

Pachur, T., M. Schulte-Mecklenbeck, R. O. Murphy, and R. Hertwig (2018). Prospect theory reflects selective allocation of attention. *Journal of Experimental Psychology: General 147*(2), 147.

Padoa-Schioppa, C. and J. A. Assad (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature 441*(7090), 223–226.

Paneri, S. and G. G. Gregoriou (2017). Top-down control of visual attention by the prefrontal cortex. functional specialization and long-range interactions. *Frontiers in neuroscience 11*, 545.

Peirce, J. W. (2007). Psychopy—psychophysics software in python. *Journal of neuroscience methods 162*(1-2), 8–13.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision 10*(4), 437–442.

Polanía, R., I. Krajbich, M. Grueschow, and C. C. Ruff (2014). Neural oscillations and synchronization differentially support evidence accumulation in perceptual and value-based decision making. *Neuron 82*(3), 709–720.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology 32*(1), 3–25.

Purcell, B. A., J. D. Schall, G. D. Logan, and T. J. Palmeri (2012). From salience to saccades: multiple-alternative gated stochastic accumulator model of visual search. *Journal of Neuroscience 32*(10), 3433–3446.

Rangel, A., C. Camerer, and P. R. Montague (2008). A framework for studying the neurobiology of value-based decision making. *Nature reviews neuroscience 9*(7), 545–556.

Rangel, A. and J. A. Clithero (2014). The computation of stimulus values in simple choice. In *Neuroeconomics*, pp. 125–148. Elsevier.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review 85*(2), 59.

Ratcliff, R. and G. McKoon (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation 20*(4), 873–922.

Ratcliff, R. and J. N. Rouder (1998). Modeling response times for two-choice decisions. *Psychological science 9*(5), 347–356.

Ratcliff, R. and P. L. Smith (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological review 111*(2), 333.

Ratcliff, R., P. L. Smith, S. D. Brown, and G. McKoon (2016). Diffusion decision model: Current issues and history. *Trends in cognitive sciences 20*(4), 260–281.

Ratcliff, R., A. Thapar, and G. McKoon (2001). The effects of aging on reaction time in a signal detection task. *Psychology and aging 16*(2), 323.

Ratcliff, R., C. Voskuilen, and A. Teodorescu (2018). Modeling 2-alternative forced-choice tasks: Accounting for both magnitude and difference effects. *Cognitive psychology 103*, 1–22.

Reeck, C., D. Wall, and E. J. Johnson (2017). Search predicts and changes patience in intertemporal choice. *Proceedings of the National Academy of Sciences 114*(45), 11890–11895.

Reynolds, J. H., T. Pasternak, and R. Desimone (2000). Attention increases sensitivity of v4 neurons. *Neuron 26*(3), 703–714.

Rigoli, F., C. Mathys, K. J. Friston, and R. J. Dolan (2017). A unifying bayesian account of contextual effects in value-based choice. *PLoS computational biology 13*(10), e1005769.

Roe, R. M., J. R. Busemeyer, and J. T. Townsend (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological review 108*(2), 370.

Roitman, J. D. and M. N. Shadlen (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of neuroscience 22*(21), 9475–9489.

Rossi, A. F., L. Pessoa, R. Desimone, and L. G. Ungerleider (2009). The prefrontal cortex and the executive control of attention. *Experimental brain research 192*(3), 489.

Sapountzis, P. and G. G. Gregoriou (2018). Neural signatures of attention: Insights from decoding population activity patterns. *Front Biosci 23*, 221–246.

Schonberg, T., A. Bakkour, A. M. Hover, J. A. Mumford, L. Nagar, J. Perez, and R. A. Poldrack (2014). Changing value through cued approach: an automatic mechanism of behavior change. *Nature neuroscience 17*(4), 625.

Scott, B. B., C. M. Constantinople, A. Akrami, T. D. Hanks, C. D. Brody, and D. W. Tank (2017). Fronto-parietal cortical circuits encode accumulated evidence with a diversity of timescales. *Neuron 95*(2), 385–398.

Shadlen, M. N. and D. Shohamy (2016). Decision making and sequential sampling from memory. *Neuron 90*(5), 927–939.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory & cognition 21*(4), 546–556.

Shafir, S., T. A. Waite, and B. H. Smith (2002). Context-dependent violations of rational choice in honeybees (apis mellifera) and gray jays (perisoreus canadensis). *Behavioral Ecology and Sociobiology 51*(2), 180–187.

Shimojo, S., C. Simion, E. Shimojo, and C. Scheier (2003). Gaze bias both reflects and influences preference. *Nature neuroscience 6*(12), 1317–1322.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of consumer research 16*(2), 158–174.

Slaughter, J. E., E. F. Sinar, and S. Highhouse (1999). Decoy effects and attribute-level inferences. *Journal of applied psychology 84*(5), 823.

Smith, E. H., G. Horga, M. J. Yates, C. B. Mikell, G. P. Banks, Y. J. Pathak, C. A. Schevon, G. M. McKhann, B. Y. Hayden, M. M. Botvinick, et al. (2019). Widespread temporal coding of cognitive control in the human prefrontal cortex. *Nature neuroscience*, 1–9.

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of mathematical psychology 44*(3), 408–463.

Smith, S. M. and I. Krajbich (2019). Gaze amplifies value in decision making. *Psychological science 30*(1), 116–128.

Smith, S. M., I. Krajbich, and R. Webb (2019). Estimating the dynamic role of attention via random utility. *Journal of the Economic Science Association 5*(1), 97–111.

Soltani, A., B. De Martino, and C. Camerer (2012). A range-normalization model of context-dependent choice: a new model and evidence. *PLoS computational biology 8*(7).

Squire, R. F., B. Noudoost, R. J. Schafer, and T. Moore (2013). Prefrontal contributions to visual selective attention. *Annual review of neuroscience 36*, 451–466.

Srivastava, V., S. F. Feng, J. D. Cohen, N. E. Leonard, and A. Shenhav (2017). A martingale analysis of first passage times of time-dependent wiener diffusion models. *Journal of mathematical psychology 77*, 94–110.

Stewart, N., F. Hermens, and W. J. Matthews (2016). Eye movements in risky choice. *Journal of Behavioral Decision Making 29*(2-3), 116–136.

Sutton, R. S., A. G. Barto, et al. (1998). *Introduction to reinforcement learning*, Volume 135. MIT press Cambridge.

Tavares, G., P. Perona, and A. Rangel (2017). The attentional drift diffusion model of simple perceptual decision-making. *Frontiers in neuroscience 11*, 468.

Thomas, A. W., F. Molter, I. Krajbich, H. R. Heekeren, and P. N. Mohr (2019). Gaze bias differences capture individual choice behaviour. *Nature human behaviour 3*(6), 625–635.

Thompson, K. G., N. P. Bichot, and J. D. Schall (1997). Dissociation of visual discrimination from saccade programming in macaque frontal eye field. *Journal of neurophysiology 77*(2), 1046–1050.

Tom, S. M., C. R. Fox, C. Trepel, and R. A. Poldrack (2007). The neural basis of loss aversion in decision-making under risk. *Science 315*(5811), 515–518.

Towal, R. B., M. Mormann, and C. Koch (2013). Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proceedings of the National Academy of Sciences 110*(40), E3858–E3867.

Treue, S. and J. C. M. Trujillo (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature 399*(6736), 575–579.

Trueblood, J. S., S. D. Brown, A. Heathcote, and J. R. Busemeyer (2013). Not just for consumers: Context effects are fundamental to decision making. *Psychological science 24*(6), 901–908.

Tsetsos, K., N. Chater, and M. Usher (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences 109*(24), 9659–9664.

Tsetsos, K., M. Usher, and N. Chater (2010). Preference reversal in multiattribute choice. *Psychological review 117*(4), 1275.

Tsetsos, K., M. Usher, and J. L. McClelland (2011). Testing multi-alternative decision models with non-stationary evidence. *Frontiers in neuroscience 5*, 63.

Turner, B. M., J. Gao, S. Koenig, D. Palfy, and J. L. McClelland (2017). The dynamics of multimodal integration: The averaging diffusion model. *Psychonomic bulletin & review 24*(6), 1819–1843.

Turner, B. M., D. R. Schley, C. Muller, and K. Tsetsos (2018). Competing theories of multialternative, multiattribute preferential choice. *Psychological review 125*(3), 329.

Turner, B. M. and P. B. Sederberg (2012). Approximate bayesian computation with differential evolution. *Journal of Mathematical Psychology 56*(5), 375–385.

Turner, B. M. and P. B. Sederberg (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic bulletin & review 21*(2), 227–250.

Tusche, A., S. Bode, and J.-D. Haynes (2010). Neural responses to unattended products predict later consumer choices. *Journal of neuroscience 30*(23), 8024–8031.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological review 79*(4), 281.

Tversky, A. and D. Kahneman (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–291.

Tversky, A. and D. Kahneman (1991). Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics 106*(4), 1039–1061.

Tversky, A. and I. Simonson (1993). Context-dependent preferences. *Management science 39*(10), 1179–1189.

Usher, M. and J. L. McClelland (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review 108*(3), 550.

Usher, M. and J. L. McClelland (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological review 111*(3), 757.

Vaidya, A. R. and L. K. Fellows (2015). Testing necessary regional frontal contributions to value assessment and fixation-based updating. *Nature communications 6*, 10120.

Vandekerckhove, J., F. Tuerlinckx, and M. D. Lee (2011). Hierarchical diffusion models for two-choice response times. *Psychological methods 16*(1), 44.

Von Neumann, J. and O. Morgenstern (1947). Theory of games and economic behavior, 2nd rev.

Voss, A., V. Lerche, U. Mertens, and J. Voss (2019). Sequential sampling models with variable boundaries and non-normal noise: A comparison of six models. *Psychonomic bulletin & review 26*(3), 813–832.

Voss, A., K. Rothermund, and J. Voss (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & cognition 32*(7), 1206–1220.

Wabersich, D. and J. Vandekerckhove (2014). Extending jags: A tutorial on adding custom distributions to jags (with a diffusion model example). *Behavior research methods 46*(1), 15–28.

Webb, R. (2019). The (neural) dynamics of stochastic choice. *Management Science 65*(1), 230–255.

White, C. N., R. Ratcliff, M. W. Vasey, and G. McKoon (2010). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology 54*(1), 39–52.

Wiecki, T. V., I. Sofer, and M. J. Frank (2013). Hddm: hierarchical bayesian estimation of the drift-diffusion model in python. *Frontiers in neuroinformatics 7*, 14.

Yang, S. and M. Lynn (2014). More evidence challenging the robustness and usefulness of the attraction effect. *Journal of Marketing Research 51*(4), 508–513.

Zhou, Y. and D. J. Freedman (2019). Posterior parietal cortex plays a causal role in perceptual and categorical decisions. *Science 365*(6449), 180–185.

Appendix A: Study 1

A mechanistic foundation of the role of visual-attention in the framing effect

Gaia Lombardi, Todd Hare, Ernst Fehr

University of Zürich

Department of Economics

Zürich Center for Neuroeconomics (ZNE)

A.1 Abstract

Subjective preferences have been shown to change across different contexts even when the options are kept constant. The framing effect is a well documented example of preference reversal in which individuals seem to change their subjective valuation of the options in the choice set. However, the neurological and internal mechanisms that leads people to change their choices depending on the different frames are still unclear. Why should mere changes in the description of an objectively identical relationship between choices and outcomes affect decision-making? Here, we examine the hypothesis that changes in framing cause changes in the allocation of visual-attention to the different options, and in turns visual-attentional changes give rise to changes in the decision process. We document that in decision making under risk the framing of sure alternatives as a gain – as opposed to a loss – induces a visual-attentional advantage for the sure option relative to the risky one which is accompanied with an increase in the choice probability of the sure option, i.e., an increase in risk aversion. Based on an evidence accumulation process (aDDM), we propose an explanation for the framing effect that is exclusively dependent on a reallocation of visual-attention through changes in the parameters of the model that do not involve a changes in the valuation of the options in the choice set. We tested our hypotheses with a hierarchical Bayesian modelling approach (HaDDM), and find that frames have an impact primarily on the initial bias of the evidence accumulation process that, combined with

changes in the allocation of visual-attention, can fully explain the observed changes in risk aversion. Our results suggest that the main drivers of framing effects in decision-making are shifts in attention allocation that alter the inner mechanisms of the evidence accumulation process.

A.2 Introduction

Subjective preferences have been shown to change across different contexts even when the options are kept constant. For example, Kahneman and Tversky have shown that humans tend to prefer sure options over risky prospects, when these options are presented as gains, but this aversion to risk is diminished if the sure options are presented as losses. This effect of framing on choices occurs even when both decision problems are identical in terms of possible outcomes (Kahneman and Tversky, A., 1979; Tversky and Kahneman, 1981). The framing effect is a well documented example of preference reversal in which individuals seem to change their subjective valuation of the options in the choice set. However, the neurological and internal mechanisms that leads people to change their choices depending on the different frames are still unclear. Why should mere changes in the description of an objectively identical relationship between choices and outcomes affect decision-making?

Here, we examine the hypothesis that changes in framing cause changes in the allocation of visual-attention to the different options, and in turns visual-attentional changes give rise to changes in the decision process. To build mechanistic and computational supports for this thesis, we adopt an attentional drift diffusion framework (i.e. an attentional drift diffusion model or aDDM). According to this model, visually guided attention temporarily biases the decision process in favour of the alternative that is attended (Ashby et al., 2016; Cavanagh et al., 2014; Konovalov and Krajbich, 2016; Kovach et al., 2014; Krajbich and Rangel, 2011; Krajbich et al., 2010; Lim et al., 2011; Schonberg et al., 2014; Shimojo et al., 2003; Stewart et al., 2016; Towal et al., 2013; Vaidya and Fellows, 2015). The advantage of using a particular framework such as the aDDM can be found in the fact that it provides accurate and quantitative predictions that can be tested in our experimental data. In particular, the aDDM makes specific predictions about the relationship between options' values, visual fixations, reaction time and choices. The key aim of this study is to adopt such quantitative predictions to make inferences about how and whether changes in preferences are influenced by visual-guided attention in the context of the framing effect. We document that in decision making under risk the framing of sure alternatives as a gain – as opposed to a loss – induces a visual-attentional advantage for the sure option relative to the risky one which is accompanied with an increase in the choice probability of the sure option, i.e., an increase in risk aversion. Based on the aDDM framework, we propose an explanation for the framing effect that is exclusively dependent on a reallocation of visual-attention through changes in the parameters of the model that do not involve a changes in the valuation of the options in the choice set. We tested our hypotheses with a hierarchical Bayesian modelling approach (HaDDM), and find that frames have an impact primarily on the initial bias of the evidence accumulation process that, combined with changes in the allocation of visual-attention, can fully explain the observed changes in risk aversion. Our results suggest that the main drivers of framing effects in decision-making are shifts in attention allocation that alter the inner mechanisms of the evidence accumulation process.

A.3 Results

Experimental Design

The main purpose of our study is to examine the attentional processes that are potentially underlying the framing effect in risky choices. Subjects in our study participated in a binary choice task in which they had to decide between a gamble and a sure option in differently framed decision trials while we simultaneously recorded their gaze patterns with eye-tracker. The sure alternative on the decision screen was either framed as a gain or as a loss as in De Martino et al. 2006). Each trial started with a screen that showed the subject's initial monetary endowment for that trial. Then, a decision screen that showed the gamble and the sure option appeared. In each trial, the expected value of the sure and the risky option was identical such that – given the natural tendency towards risk aversion – subjects were expected to show an overall preference towards the sure option but in the presence of a framing effect this tendency should be magnified in the gain condition and mitigated – or even overturned in favour of risk seeking behavior – in the loss condition.

As shown in figure A.1, each option was presented to the subjects as a combination of a rectangle and a pie-chart. The color-filled area of the rectangle represented the part of the initial endowment that the participants could keep or lose, whereas the color-filled area of the pie-chart denoted the probability with which money could be kept or lost. The color with which the shapes were filled indicated whether money could be kept or lost. For example, in the two trials shown in figure A.1 subjects received in both conditions an initial amount of money of 100 Swiss Francs that was indicated in the middle of the screen. Then, they had to look at this amount for at least 2 seconds in order to move on to the decision screen; this feature of our design ensured that their initial gaze on the decision screen would not be biased towards any of the two options.

In the gain condition of figure A.1 subjects had to decide between a risky option on the left-hand side of the screen that offered them the possibility to keep the 100 Swiss Francs with 70% probability and a sure option on the right-hand side of the screen that offered them to keep 70 out of the 100 Swiss Francs with 100% probability. In the loss condition, the final consequences of each option were exactly the same as in the gain condition, but the sure option was framed as a loss, i.e., they could lose 30 out of the 100 Swiss Francs with 100% probability. In figure A.1, the blue colors represented the 'keep' frame, and orange indicated the 'lose' frame. The meaning of the colors was randomized between subjects, and the location of the sure and risky option on the screen was randomized between trials and subjects. It is worth mentioning that during the presentation of the decision screen, subjects were not shown the numbers and words; they could only see shapes and colors. This feature of our design allowed us to have longer fixation duration and minimize the attentional component that might come from memory – i.e. it is much easier to remember something written in number and letters than a perceptual component of the task. However, we ensured that subjects knew the meaning of the shapes and colors with two extensive training phases before we started the main experiment.

In the first training phase, subjects were presented with many subsequent pie-charts and they had to indicate the correct probabilities represented by the color-filled area in the pie-charts. In addition, they had to specify whether the filling colors indicated a loss or a gain in each trial. The number of pie-charts presented (i.e. trials) to each subject depended on the her own performance with a minimum of 30 presentations. Thus, after indicating the correct probabilities of 30 pie-charts for which they received feedback on correct/incorrect responses, the subject had to respond correctly to 10 pie-charts in a row in order to end the task.

In the second training phase, subjects were presented with many subsequent rectangles and they had to indicate the correct amount of money represented by the color-filled areas and the meaning (gain or loss) of the colors. Similarly to the real task, before each presentation a monetary endowment was associated with each rectangle, and the amount of money represented in the color-filled area of the rectangle depended on this endowment – i.e the color-filled area was always a proportion of the initial endowment. As for the first training phase, the number of trials was performance dependent on each subject with a minimum of 50 presentations. Thus, after indicating the correct amount of money of 50 rectangles and the meaning of the colors - they received feedback on correct/incorrect responses on each trial -, the subject had to respond correctly to 10 rectangles in a row in order to end the task (for more details see Methods section).

In addition to the training phases and the choice task, our subjects performed a second task – the certainty equivalent elicitation task. In each trial of this task they were asked to choose between a gamble and many possible sure amounts of money presented as a list (figure A.2; also, see details in Methods section). This task provided us with a method to elicit the certainty equivalent (CE) for each gamble presented in the preceding choice task - i.e. the guaranteed amounts of money that each subject would view as equally desirable as the gambles. The elicited certainty equivalents (CEs) served as an important input for the estimation of the parameters of an attention-based drift-diffusion model (aDDM) described in more detail below.

Choice behavior

Similar to De Martino et al. (2006) subjects chose the sure option on average significantly more often in the gain condition than in the loss condition (Fig A.3a). The data also show that there is an overall tendency to choose the sure option more than the gamble in both conditions but this tendency is much more pronounced in the gain condition. The data clearly indicate a behavioral framing effect at the individual level because almost all subjects chose the sure option more frequently in the gain condition (Fig A.3b). This result is further corroborated by a logistic mixed effects regression with random effects (Table A.1, model 1) that shows that the loss condition significantly decreases the probability of choosing the sure option (beta = -0.93, $p < 0.001$), and this reduction in the choice probability of the sure option is larger at higher expected values of the options (beta = -0.26, $p = 0.047$). Furthermore, if the subjective attractiveness of the sure option relative to the gamble – measured by the difference between the monetary value M_s of the sure option and the certainty equivalent CE of the gamble – increases, the probability of choosing the sure option increases (beta = 0.62, $p = 0.009$). Finally, model 1 in Table A.1 also answers the question how choice behaviour responds to a general rise in the stake level, i.e., a rise in M_s =Ev: at higher stakes subjects increase their preference for the sure option both in the gain and the loss condition (beta = 0.63, $p < 0.001$).

Model Framework

Because we measure subjects' gaze with the eye tracker we can use patterns of visual-fixations to provide a deeper mechanistic understanding of the role of framing in subjects' decision process. In particular, the aim of this paper is to investigate the role of visual-attention on the framing effect. Namely, we want to study how visual-attention is related to differences in choices and reaction time across frames and how it interacts with the values of the options on the screen. For this purpose, we use the attentional drift diffusion model (aDDM) to derive testable predictions about the relationship between subjects' visual fixation patterns and their choices in the decision task. The rationale for adopting this model instead of other attentional based models is noteworthy and need some careful explanation. We will first present the aDDM in details, and then illustrate the features and the assumptions that make this model suitable for our data set and for investigating the role of visual-attention in the framing effect.

The aDDM assumes that the brain, over time, computes and updates a relative decision value (RDV) signal, which depends on the difference between the noisy accumulated evidence for the two decision options. The speed at which the RDV moves in favour of one of the two options is proportional to the subjective value difference of the options multiplied by a drift parameter that determines the slope of the evidence accumulation. A key feature of this model is that the evidence accumulation process depends on where the subject is looking: on average, an individual accumulates more evidence in favour of the attended option than the unattended one. This accumulation mechanism has important implications for the decision process. In particular, the visual fixation pattern is assumed to influence choices, i.e., the more time a subject spends at looking at one option, the more likely she is to choose that option. The aDDM can be formalized as follow:

$$
\begin{aligned}
& dV(t) = \mu(t)dt + \sigma dW(t), \\
& V_0 = x_0, \\
& \tau = \inf\{t > 0 \mid V(t) \notin (B, -B)\},
\end{aligned} \tag{A.1}
$$

where $V(t)$ is the evidence accumulated at time t (or relative decision value), τ is the first passage time (or reaction time) when the evidence crosses a certain threshold B, $\mu(t)$ is the drift rate or speed of the accumulation process, σ is the constant diffusion rate or the standard deviation of the brownian motion and x_0 the initial bias parameter that is positive when there is an initial bias towards the sure option and negative if there is an initial bias towards the gamble. x_0 is a bias towards one or the other alternative that is independent from the evidence accumulation process, i.e., it could arise from features of the decision task that trigger a spontaneous attraction or aversion towards one option. For example, if there is a spontaneous aversion against accepting a sure loss, then x_0 may be negative. The evidence $V(t)$ evolves in time according to a biased random walk with Gaussian increments, i.e. $dV(t) \sim N(\mu(t)dt, \sigma^2 dt)$, where $\mu(t)$ depends on which option is attended at time t as follows

$$
\mu(t) = \begin{cases} \delta(M_s - \theta CE) & \text{if the sure option is attended} \\ \delta(\theta M_s - CE) & \text{if the gamble is attended} \end{cases} \tag{A.2}
$$

where M_s is the monetary amount offered in the sure option, CE the certainty equivalent of the gamble, δ is the constant drift parameter that controls the speed or slope of the evidence accumulation, θ is the attentional discount factor that is always between 0 and 1. Because the unattended option is multiplied by θ and obeys the restriction $0 < \theta < 1$, the model captures the idea that less evidence is accumulated for the unattended option. Previous research has indicated that the discount factor is typically considerably smaller than 1 for virtually all subjects (Krajbich and Rangel, 2011; Krajbich et al., 2010; Lim et al., 2011).

Let us consider now limitations, assumptions and advantages of adopting this model. In general, the standard DDM – the version without the visual-attention component (i.e. $\theta = 1$) – predicts a specific relation between choices and reaction time. In particular, it predicts a decrease in reaction time as the absolute subjective value difference between the two options, $|M_s - CE|$, increases. In other words, the easier the choice the faster the response should be. In addition, the model predicts that the reaction time distribution has the shape of an exponentially modified Gaussian (ex-Gaussian) distribution. These two basic predictions should be satisfied in the data to have trust in the applicability of the model to our data. It turns out that our data meet both of these predictions. Table A.1 in the appendix regresses the logarithm of the response time (RT) on (i) the loss condition, (ii) the absolute subjective value difference $|M_s - CE|$, and (iii) the interaction of both variables. This regression shows that higher absolute value differences are associated with lower RTs. Figure SA1 in the supplementary materials shows that the RT-distribution is right skewed with the characteristic long right tail indicated by an ex-Gaussian.

Considering the aDDM, an assumption of the model that might be considered a weak limitation of this framework is the randomness of visual gazes. Clearly, the model assumes that gaze patterns do not depend on options' values nor on attributes' values, and the aDDM does not make any prediction nor inference on the transition probability of a fixation from one option (or attribute) to another. Thus, visual search strategies or probabilities of gazes to the attributes are not accounted for in this model. However, we will argue that this assumption – or better, this non-assumption – of the model should not affect our interpretation of the data and our results of the model fitting.

The first attempt to overcome this weakness is through the way we formalized the model in equation A.1. A fixation is considered belonging to a specific option whenever one of its attributes is attended, i.e. any fixation that fell into the money or the probability attributes was consider a fixation to the option to which the attributes belong. This allows us to have random fixations from one option to the other as there are only two options in the choice set. In other words , in the model we formalized, equation A.1, the money and the probability attributes are integrated together to form options' values that are then compared.

However, people often make choices by comparing the options' attributes (e.g., Gonzalez-Vallejo, 2002) instead of the integrated values. We try to investigate whether we can use the assumption that our participants are mainly integrating the attributes of the options. Thus, we analysed gaze transitions between and within the options' attributes in our data to test whether subjects were performing more an attribute comparison or an attribute integration type of strategy. We based our investigation on the premise that participants who look more within options (i.e., between attributes of the same option) are more likely to integrate the attributes of an option, whereas participants that look more between different options' attributes of the same type (i.e., between money attributes or between probability attributes of different options) are more likely to compare the attributes of different options instead of integrating the attributes of a given option as in Reeck et al. 2017. Figure SA2 in the supplementary materials shows that on average people make much more within option transitions than between option transitions, suggesting that they are more involved in integrating the options' attributes. This is also supported by an individual level analysis of within and between option transitions. Overall, 27 out of 30 subjects showed gaze patterns that indicate a prevalence of attribute integration over attribute comparison (supplementary material table SA4).

These findings are evidence for the view that using the integrated values of the options should not influence the results, and that the aDDM assumptions are not violated in our data set. Thus, we can apply this model to our data and also take advantage of the fact that it makes strong testable predictions about the link between visual fixation patterns and choices, as well as the role that the discount factor plays in this process.

Predictions on the role of visual-fixations for choice behavior

In the aDDM visually guided attention plays a causal role in individuals' choice behavior: the longer one option in the choice set is visually fixated in a trial the more likely it will be chosen in that trial. Thus, if the framing of options affects subjects' visual-attention pattern the model predicts that the change in the attention patterns across frames should be associated with a change in choice frequencies across frames. In addition, the aDDM predicts an interaction between the values of the options and the degree to which visual-attention influences the evidence accumulation process; this interaction follows from the multiplicative components (θ CE and θM_s) in formula A.1, as opposed to other attentional models like the decision by sampling (Noguchi and Stewart 2018) or the additive model (Cavanagh et al. 2014). Finally, yet importantly, the model makes clear predictions about reaction time in relation with fixations and option values.

Thus, several specific predictions about the relationship between choice behavior, visual-attention, reaction time, the option values and their interactions can be directly derived from the model by assuming a discount factor strictly smaller than 1.

First, the model predicts that if, ceteris paribus, the sure option has a bigger fixation time advantage over the risky option in the gain condition compared to the loss condition the sure option also has a bigger choice advantage in the gain relative to the loss condition. Thus, a larger advantage in fixation time towards the sure option in the gain condition relative to the loss condition could by itself explain the framing effect in choices. The intuition for this prediction is the following. If one option receives more overall attention in a trial, more evidence is accumulated in favour of this option relative to the unattended option: the attended option gains an advantage in evidence accumulation that is proportional to $1-\theta > 0$ and therefore it reaches the decision threshold faster than the alternative option. As a consequence, the longer attended option has a higher probability of being chosen.

Second, the above prediction should not only hold on average but it should also hold at the individual level. Subjects that exhibit a greater change in the fixation time advantage for one option across frame conditions should also display a greater change in the choice advantage for that option across conditions. The intuition for this hypothesis is that an individual with a higher fixation time advantage towards one option relative to the other accumulates more evidence in favour of the advantaged option relative to an individual that has a smaller or no fixation time advantage for that option.

Third, the intuition described in the first prediction should also hold at the level of individual trials, regardless of the framing condition: in choice trials in which an option is systematically advantaged in terms of fixation time, this option should have a higher probability of being chosen relative to the choice trials in which the other option has an advantage in fixation time. With regard to our experiment this means that for a given framing condition – regardless of whether it is the loss or the gain condition – the sure option should be chosen more often in those trials in which it has a fixation time advantage over the risky option compared to trials in which the fixation time advantage is in favour of the risky option.

Fourth, another frame-independent prediction of the aDDM is that a *given* fixation time advantage for one of the options translates into a larger choice advantage for that option at higher values of *both* M_s *and* CE when (M_s - CE) is kept constant. Intuitively, this follows from the fact that at higher values of both M_s and CE the attentional discount $(1-\theta)M_s$ and $(1-\theta)$CE is higher. Yet, the aggregate effect of the attentional discount on choices depends on which option has a fixation time advantage. If, e.g., the subject looks more often at the sure option in a trial the discount θ CE is more often relevant compared to the discount θ M_s. In other words, a given fixation time advantage of the sure option leads to a larger choice advantage for that option at higher overall values for M_s *and* CE.

Fifth, the aDDM also generates a hypothesis regarding the *interaction* between the values of the two options and the fixation time. The model predicts that the higher the value of *both* options the higher the influence of a *change* in fixation time on choices. The source of this interaction effect is again the higher attentional discounting ($(1-\theta)M_s$ and

$(1-\theta)$CE) that occurs if both options have a higher value. Therefore, a *rise* in the fixation time of, say, the sure option over the gamble leads to a stronger choice advantage of the sure option if both options have a higher value.

In addition to the choice-related predictions, the aDDM also generates hypotheses about how fixation time advantages for one option affect the overall reaction time. This allows us to formulate two further predictions concerning the reaction time.

First, the stronger the fixation time advantage of one option the faster a choice should be made because the non-attended option will be attentionally discounted for a larger proportion of time (see simulated example in figure A.4). To illustrate this with regard to our experiment, recall from the first hypothesis above that if, say, the sure option has a larger fixation time advantage in the gain condition relative to the loss condition it is predicted to have a larger choice advantage in the gain relative to the loss condition. Yet, a larger fixation time advantage for the sure option in the gain condition also means that sufficient evidence in favour of this option is accumulated more quickly in the gain condition such that the sure option is chosen more quickly in the gain compared to the loss condition.

Finally, if we observe a general fixation time advantage for the sure option then we should also observe faster response times in choosing the sure option compared to choosing the gamble.

Results on the role of visual-fixations for behavior in the binary choice task

All of the above hypotheses predict a systematic relationship between the visual-attention patterns, subjects' choices', subjective values and value differences (M_s - CE), and response times. Our first prediction above concerns the relation between choices and fixation patterns. The aDDM predicts that if the gain condition affects subjects' visual-attention in favour of the sure option, the framing effect on choices can be at least partly explained as an attention-driven phenomenon. However, do the gain and the loss frames really affect visual-attention in this way? We indeed find that there is a strong and statistically significant advantage in the allocation of fixation time to the sure option in the gain condition that is significantly reduced when subjects are in the loss condition (Fig A.5A and Table A.2). Figure A.5A shows that at all levels of the expected value of the gamble, Ev, the fixation time advantage of the sure option is higher in the gain compared to the loss condition. The regression in Table A.2 column 2 indicates that the fixation time advantage of the sure option is reduced by 25 percentage points in the loss condition ($p < 0.01$). Together with the behavioural framing effect in figure A.3, this result is perfectly consistent with the first hypothesis: if there is a larger fixation time advantage for the sure option in the gain relative to the loss condition, then we should observe the same qualitative result in the domain of choice behaviour, i.e., a larger choice advantage for the sure option in the gain relative to the loss condition.

The second prediction above states that if the different frames affect choices by triggering changes in fixation time patterns then we should observe this not only at the aggregate but also at the individual level: Individuals for whom the gain condition triggers a larger fixation time change in favour of the sure option should display a larger change in the frequency of choosing the sure option in the gain relative to the loss condition. Figure A.5B supports this hypothesis. We observe a high and significant correlation (Pearson index 0.765, $p < 0.001$) between the difference in the fixation time that is allocated to the sure option in the gain condition relative to the loss condition, and the difference in the probability of choosing the sure option in the gain relative to the loss conditions. In other words, individuals that show a higher fixation time advantage towards the sure option in the gain relative to the loss condition also show a higher choice advantage towards the sure option across the two conditions.

The third hypothesis above states that irrespective of the framing condition a fixation time advantage for the sure option in a given trial leads to a higher probability of choosing the sure option compared to trials in which the gamble has a visual-attention advantage. Figure A.5C shows that this is indeed the case. In both the gain and the loss condition the probability of choosing the sure option is substantially lower in those trials in which the gamble has a fixation time advantage.

In the binary choice task the monetary value of the sure option, M_s , is always equal to the expected value of the gamble, Ev, which – due to subjects' known tendency towards risk aversion – generally favours the choice of the sure option. Figure A.3aa shows, however, that at higher expected values (M_s = Ev) of the two options the tendency to choose the sure option even increases. Moreover, this increase in risk aversion at higher stakes occurs in both the gain and the loss condition. The aDDM can explain this pattern if subjects allocate relatively more visual-attention to the sure option at higher expected values. Then the gamble suffers from more attentional discounting at higher expected values and therefore the sure option is more often chosen. Figure A.5A indeed shows that more visual-attention is allocated to the sure option both in the gain and the loss condition when the stakes (M_s = Ev) are higher, and Table A.2 shows that this increase in visual-attention in response to a rise in Ev is significant ($p < 0.001$). These findings are consistent with the idea that the increase in visual-attention towards the sure option may be a source of the increased behavioural risk aversion at higher levels of Ev.

Our fourth hypothesis above is related but not identical to the issue of increasing risk aversion. It states that a *given* fixation time advantage for one option translates into a higher choice advantage for that option at higher stakes when keeping the difference (M_s - CE) constant. The reason behind this hypothesis is that at a generally higher level of M_s and CE the visually disadvantaged option suffers for a greater attentional discount $(1-\theta)M_s$ or $(1-\theta)CE$. As in our experiment, the gamble is generally the option that receives less attention we should observe that the frequency of choosing the sure option increases as M_s increases (when controlling for $M_s - CE$). Regression 1 in Table A.1 shows that this prediction is met. The coefficient on M_s (resp. Ev) is significant in all three models in Table A.1.

Next, we examine the hypothesis (prediction five above) that there is an interaction effect between M_s and the fixation time advantage of an option. More specifically, this hypothesis states that a *rise* in the fixations towards the sure option relative to the gamble leads to a stronger choice advantage of the sure option if both options have a higher value because at the higher values of the two options the visually less attended option (i.e., the gamble) is discounted more heavily. Model 3 in Table A.1 shows that the coefficient on the interaction term between M_s and the fixations towards the sure option is positive and significant.

Finally, we investigate the effect of fixation time advantages on reaction time. In figure A.5A we have seen that the sure option enjoys a larger fixation time advantage in the gain condition relative to the loss condition and in the previous section we hypothesized that this should translate into faster decisions for the sure option in the gain condition compared to the loss condition. The same logic should apply with regard to stake size. Figure A.5A shows that a rise in Ev is associated with a large rise in the attention towards the sure option. According to the aDDM this implies that the sure option should be chosen more quickly at higher levels of Ev. Figure A.6b shows that both of these predictions are met by the data: the sure option is chosen more quickly in the gain condition and at higher levels of Ev and this latter effect holds both in the gain and the loss condition. Further evidence for these results is provided in the supplementary materials (Table SA2) where we regress the log of response times on the loss condition (i.e., omitted category = gain condition), M_s =Ev, and the interaction of the two. The regression shows that the RTs are significantly larger in the loss condition and significantly faster at higher levels of M_s . Furthermore, we know from figure A.5A and Table A.2 (column 2) that there is a fixation advantage for the sure option that is larger at higher stakes (M_s =Ev). Therefore, the model implies that there should be a faster response time when the the sure option is chosen compared to when the gamble is chosen at higher M_s =Ev. Figure A.6a shows that this is indeed the case and regression 2 in Table SA2 in the supplementary materials corroborates this prediction econometrically. The regression shows a significantly positive interaction between "Gamble chosen" and M_s on response times.

The previous evidence provides support for the view that the framing effect is an attention-driven phenomenon. It appears that the gain frame directs more attention towards the sure option which then increases the probability of choosing the sure option more frequently and more quickly. However, how do we know that the link goes from visual-attention to choices and not the other way around? In other words, how do we know whether individuals choose what they look or whether they look at what they prefer? Fortunately, our data enable us to directly examine this question by looking at the correlation between visual-attention and the options' values as well as the subjective value differences. We found that both middle fixation duration and first fixation duration do not correlate with the options' values, M_s and CE, respectively. Figure A.7 shows the average duration of middle and first fixations as a function of options' values. In this figure the options' values are defined as CE if the attended option is the gamble or M_s if the attended option is the sure option. The figure shows that options with higher values do not attract longer fixation duration. We also run two separate linear mixed-effect regressions of the logarithm of middle fixation duration and first fixation duration on the options' values and did not observe a significant correlation (see table SA7 in the supplementary materials and figure A.7, beta = 0.022, p-value = 0.11 and beta = 0.028, p-value = 0.72, respectively). Further, the regressions in Table A.2 show that neither the percentage of fixations to the sure option nor the fixation time advantage of the sure option are affected by the relative subjective attractiveness (M_s-CE) of the sure option. In both regressions these coefficients are rather small and clearly insignificant (p = 0.24 and p = 0.52 in regressions 1 and 2, respectively). Thus, our data do not support the view that the attractiveness of the options influences visual-attention. Rather, they are consistent with the view that the gain frame changes people's visual-attention in favour of the sure option and that this change induced an increase in the choice frequency of this option. However, there is still some unexplained variance in our data that cannot be fully explained by visual-attention alone. In fact, despite accounting for the proportion of fixation time towards the sure option in the logistic mixed-effect regression model (Table A.1, model 2), there is still a main effect of the loss condition in which the probability of choosing the sure option is significantly decreased (beta=-0.89 , p-value<1e-5). In addition, there is a significant negative interaction between the loss condition and the proportion of fixation time towards the sure option (beta = -1.63, p-value = 0.01), suggesting that the impact of fixation time towards the sure option on the probability of choosing the sure option is reduced in the loss condition. This interaction effect between visual-attention and the loss condition means that, for a given visual fixation advantage of the sure option, the loss condition transforms this advantage to a lower degree into a choice advantage for the sure option. But what does that exactly mean in terms of the underlying mechanisms that link visual-attention and choices? Additionally, are there other aspects and mechanisms in the evidence accumulation process that play a role in the framing effect? For instance, is the discount factor or the drift rate affected by the frame? Or is there an initial bias in the accumulation process driven by the frame? In our attempt to answer these questions, we fitted the aDDM to our data first separately for gain and loss conditions.

Towards a mechanistic understanding of the framing effect

As previously discussed, the main features of the aDDM are the drift rate parameter (δ) that determines the speed of the evidence accumulation, the discount factor (θ) that "penalizes" the unattended option, the initial bias parameter (x_0) that provides one or the other option with an initial advantage in terms of evidence, and the threshold parameter (B) that determines how much evidence is needed to reach a decision. We are now interested in examining how framing affects the different parameters of the aDDM. As we already discussed in the previous section, changes in visual fixation patterns alone are unlikely to be the only source through which the framing effect is generated. On the contrary, it seems easily possible that the parameters of the aDDM vary across conditions. For example, the interaction between the loss condition and the fixation time advantage for the sure option is difficult to explain without a change in parameters across frames. In addition, quantitative evidence on how the different parameters change across frames might – in combination with the fixation time changes induces by framing – make it possible to fully account for the changes in choice behaviour across frames.

To identify changes in parameters between frames we fitted the aDDM model to our data with a hierarchical Bayesian method (HaDDM). The hierarchical estimation of the mean population parameters at the group level allows us to make inference about the parameters accounting for variation coming from sources like individual differences. Thus, we can benefit from this method to investigate the group level differences in parameters across gain and loss fitting the model separately for the two conditions (Model fitting of the Methods section).

We show the posterior distributions of the estimated parameters of the aDDM model in figures A.8 – A.11. Each of these figures has three elements – A, B and C. In Figures A.8A – A.11A we show the posterior distributions of the parameter means – for each aDDM parameter – at the Group level. The Group level distribution for the gain frame is in blue while for the loss frame it is illustrated in red. These results suggest that framing changes primarily the initial bias and in the drift rate parameter – both of which are higher in the gain condition – but the gain frame also slightly lowers the attentional discount parameter, i.e., there is a higher attentional discount for the unattended option in the gain condition compared to the loss condition. In addition, the variance of the attentional discount parameter is considerably smaller in the gain condition. The figure for the threshold parameter indicates that it is slightly larger in the loss condition, i.e., subjects seem to demand more evidence until they make a choice in the loss condition.

To corroborate these results and further investigate how the parameter changes between loss and gain conditions are related to the framing effect, we analysed the parameters' means at the individual level for individuals with a large and a small behavioural framing effect. In particular, we split subjects in two groups based on whether they showed a behavioural framing effect below or above the median. Figures A.8B – A.11B upper plots show the posterior distributions of the mean parameters of the individuals that show a below-median behavioural framing effect, i.e., for them the difference in the probability of choosing the sure option between gain and loss conditions is below the median. Figures A.8B – A.11B lower plots show, in contrast, the posterior distributions of the mean parameters of the individuals that show an above-median behavioural framing effect. A Comparison of figures B between upper and lower plots for each parameter shows that the parameter changes between the frames are primarily driven by individuals with an above median framing effect. In fact, for these individuals the drift rate parameter (δ) and the initial bias parameter (x_0) are significantly larger in the gain condition ($p < 0.01$) while the attentional discount parameter (θ) is (marginally) significantly smaller in the gain condition ($p = 0.055$), i.e., the unattended option is more heavily discounted in the gain condition.

How do these differences in parameters help us understand the behavioural framing effect and the associated response time patterns? Clearly, an increase in the drift rate, keeping all the other parameters fixed, will affect the evidence accumulation process resulting in a reduction in reaction time and an increased frequency of choosing the option with a fixation time advantage. Thus, because the sure option generally enjoys a fixation time advantage, the higher drift rate in the gain condition compared to the loss condition is one component in explaining why the sure option is more frequently chosen in the gain condition. In addition, the faster reaction times that we observe in the gain compared to the loss condition (see table SA2, supplementary material) can also be explained by the higher drift rate in the gain condition.

As previously discussed, the initial bias parameter captures some a priori advantage at the very beginning of the accumulation process – an advantage that may be due to individuals' spontaneous response when they notice they face the gain or the loss condition. In our data, we measure a significantly positive initial bias in the gain condition towards the sure option that is reduced to zero in the loss condition, which further adds to the higher choice frequency for the sure option in the gain relative to the loss condition.

A possible explanation for this finding could be due to the way the options are presented in our experiment. Two aspects of the binary choice task may capture participants' attention before they start evaluating and comparing the available options. The first one is the fully coloured pie-chart that is always associated with the sure option, which makes the sure option immediately recognisable before knowing the value of the options. Secondly, the colour associated with gain and loss does not require a valuation of the options and is instantly noticeable as soon as the decision trial appears on the screen. Thus, a plausible speculation could be that subjects see instantly which option is the sure option and have some initial bias towards it. However, when the loss colour is present on the screen, i. e., they are facing a loss condition trials, this feature acts against the previous advantage for the sure option reducing the initial bias to zero. Thus, the initial bias as well may be considered an attention related mechanism for which, however, it is harder to empirically measure its link with visual-attention.

The above discussion shows that the observed changes in key parameters of the aDDM between the gain and the loss condition help us to understand the direction of the observed choice and RT changes between the conditions. But to what extend are these parameter changes – in combination with the observation fixation time changes – capable of fully explaining the behavioural framing effect? In order words, what is the precise quantitative role of the observed parameter changes in the explanation of the behavioural framing effect?

We ran a logistic mixed-effect regression with random effects for subject specific slopes and constants with the probability of choosing the sure option as the dependent variable. As explanatory variables with included the loss condition (omitted category = gain condition), the percentage of fixation time for the sure option, the subjective value difference, the monetary expected value of the options (M_s = Ev), and, in addition, for every subjects their mean value for each parameter of the aDDM for gain and for the loss condition, and all the interaction with proportion of fixation time to the sure option. Notably, the results show that the dummy variable for the loss condition has now an insignificant coefficient that is close to zero (beta = -0.18, p-value = 0.33), implying that the framing induced changes in visual-attention in combination with the induced changes of the aDDM parameters now fully account for the behavioural framing effect.

To better estimate the importance of each component we calculated the R2 for the full model (Pseudo-R^2 = 0.32) and the partial Pseudo-R^2 for each component. The proportion of fixation time to the sure option is the variable that most of all can explain the variance in the data (Pseudo-R^2 = 0.12), and although the parameter of the aDDM contribute to account for part of the framing effect their Pseudo-R^2s are very small (initial bias Pseudo-R^2 = 0.017, drift rate Pseudo-R^2= 0.005, all the parameters together Pseudo-R^2= 0.034, and condition dummy Pseudo-R^2= 0.004).

However, it is worth noticing that when comparing three different logistic mixed-effect regression models (figure A.12) for the probability of choosing the sure option, a model with proportion of fixation time and its interaction with the fitted parameters of the aDDM (model C, figure A.12) not only fits the data better than other models, but also significantly outperforms a model in which the condition is specified as a variable (dummy regressor for condition type, model B, figure A.12), (χ^2= 570.77, p-value < 2.2e-16), and a model with only proportion of fixation time towards the sure option (model A, figure A.12), (χ^2= 192.34, p-value < 3e-12). In addition, the model with only proportion of fixation time fits the data significantly better than a model with a condition dummy (χ^2= 378.43, p-value < 2.2e-16). To conclude our results suggest a prominent role of visual-fixation time in the framing effect that interacts with the decision process changing the speed of the accumulation and the initial advantage of the sure option across frames.

A.4 Discussion

Visual-attention has been shown to influence choices towards the longest attended option in value-based decision-making (Krajbich et al. 2010; Krajbich and Rangel 2011; Krajbich and Rangel 2012; Towal et al. 2013), and this bias is present even when the duration of fixations is experimentally manipulated (Shimojo et al. 2013; Armel et al. 2008; Tavares et al. 2017).

Here, we reported for the first time evidence for a bias in visual-attention allocation linked to the framing effect in a lottery choice-task, and showed that this bias in attention allocation was highly correlated with the intensity of the framing effect at the trial and at the individual level.

Furthermore, the purpose of this study was to investigate how visual-attention is related to differences in choices and reaction time across different frames and how it interacts with the values of the options in the choice set. Thus, with the aim of building mechanistic and computational supports for the hypothesis that attention has a prominent role in explaining the framing effect, we adopted an attentional drift diffusion framework and tested the quantitative predictions of the model about the interaction between choices, fixations, options' values and reaction time in our lottery data set.

We found that all the predictions of the aDDM are confirmed in our data. Specifically, we observed visual-fixations interacting consistently with choice frequency across frames and impacting reaction time accordingly to the model assumptions. Furthermore, we observed evidence for the effect of the multiplicative interaction between visual-fixations and options' values, i.e. attention seems to modulate the value of the attended option (or underweight the value of the unattended one).

These results all together provide support for the view that the framing effect is an attention-driven phenomenon. It appears that the gain frame directs more attention towards the sure option which then enhances the probability of choosing the sure option more frequently and more quickly. Nevertheless, this evidence is not sufficient for claiming a link between choice frequency and visual-attention that goes from fixations to choices and not vice-versa, i.e. decision makers choose more often what they attended to and not necessarily look more often at what they prefer. Thus, we investigated the relation between fixations duration and options' value to rule out an independence of subjective values and value differences from visual-fixation time in our task. We found that both middle fixation duration and first fixation duration do not correlate with options' values, and also we do not find a significant effect on visual-fixation time of value differences in our data. Clearly, this is not a proof of causality, a rigorous external manipulation of attention would be necessary for it, however, this is undoubtedly further validation of a main impact of visual-attention on choices independently of options' values, corroborating the hypothesis that individuals are more likely to choose what they attend.

Once all the predictions of the aDDM were tested, we could fit the a hierarchical Bayesian version of the model (HaDDM) to our experimental lottery data set to investigate how the framing affects the different aspects of the evidence accumulation process. Hierarchical Bayesian methods (Vandekerckhove, Tuerlinckx, and Lee, 2011; Wiecki, Sofer, and Frank, 2013; Wabersich and Vandekerckhove, 2014) have been shown to be superior and extremely effective in computational model fitting compared to standard methodologies such as maximum likelihood (MLE), because they allow a posteriori inference on the entire distribution of the parameters rather than just the most likely estimate, and they enable group and subject parameters to be estimated simultaneously at different hierarchical levels, thus providing robust estimates of a model's free parameters without ignoring or over-weighting individual differences. Thus, we could benefit from this method to investigate the group level differences in parameters across gain and loss fitting the model separately for the two conditions.

Our results from the fitting procedure showed mainly a change in the initial advantage, i.e. positive initial bias parameter, from choosing the sure option in the gain condition to no advantage in the loss condition, and a change in drift rate. The initial bias parameter, which capture the tendency to prefer one option before the evidence accumulation process starts, could reflect in this task how much attention the color of the options captures before the decision-maker even starts evaluating the options in the choice set. The drift rate parameter change across gain and loss could explain the mean difference in reaction time between the two conditions.

Ultimately, when running a choice model on the probability of choosing the sure option including the parameters of the aDDM and all their interactions with proportion of fixation time to the sure option, we found that we could completely explain, i.e. reduce to zero, the framing effect on choices.

Thus, our results show one possible way through which framing could change the subjective values of options, visual-attention. We were able to disentangle the effect of attention during the decision process and explain all the variability of choice difference across conditions without directly touching subjective values among gain and loss.

A.5 Methods

Subjects

Thirty-six healthy subjects (12 females) participated in our experiment. Six of them were excluded from the experiment because they failed to understand the choice task. Subjects received monetary compensation for their participation, were informed about all aspects of the experiment and gave written informed consent. The experiments conformed to the standards of the Declaration of Helsinki and the Human Subjects Committee of the University of Zürich approved the experimental protocol.

Tasks

This experiment used a modified version of the gain-loss paradigm in De Martino and Dolan 2006. The experimental design was divided into four parts: two training phases, a decision phase and an elicitation phase.

During the first training phase, subjects were trained to learn how probabilities in the decision task were represented (figure SA3a) through several trials – the number of trials depended on the performance of each participant, from a minimum of 20 until the subject performed 10 corrected probability estimations in a row. At the beginning of each trial, a computer displayed an amount of money in the upper part of the screen and a pie-chart in the middle. A fraction of this pie-chart was filled either in blue, orange or green. The other fraction was filled in gray. One colour (for instance, orange) indicated that the monetary amount displayed in the upper part of the screen was a loss, while the other colours (for instance, blue or green) indicated that this amount was a gain. The meaning of the colours was randomized among participants, but kept fixed for the duration of the entire experiment for a single subject. The size of the filled area in the pie-chart indicated the probability of keeping/losing the monetary amount. In each training trial, subjects had to indicate whether the monetary amount was a gain or a loss; and to report the exact probability represented by the colour-filled area.

In the second training phase, subjects learned to estimate monetary amounts (figure SA3b) through several trials – the number of trials depended on the performance of each participant, from a minimum of 50 until the subject performed 10 corrected money estimations in a row. Subjects started each trial of this training phase with an initial monetary endowment between 10 and 100 CHF displayed in the middle of the screen. After being informed about the endowment, a rectangular shape appeared on the screen. As before, a fraction of this rectangle was gray and the remaining was filled either in blue, orange or green. One colour (for instance, orange) indicated that the amount was a loss, while the other colours (for instance, blue or green) indicated that the amount was a gain. In each trial of this training phase, subjects had to indicate whether the amount of money was a loss or a gain, and they had to report the exact amount of money represented by the proportion of the colour-filled area of the rectangle that they could keep or lose from the initial endowment received. For example, when the filled region of the rectangle was orange, the coloured area indicated a proportion of the initial endowment to be lost. On the contrary, the filled area indicated the proportion of the initial endowment to be kept by the subject if, for example, was coloured in blue or green.

After the training sessions, subjects performed the choice task. During this decision phase (figure A.1), subjects had to make 150 decision trials between a sure option and a gamble. 70 of these trials were framed as a gain, 70 framed as loss and 10 were 'no-brainer' trials, i.e. one of the two options was clearly better in terms of money and probability than the other, which allowed to control that subjects correctly understood the task. At the beginning of each trial the participant was presented with a monetary endowment (e.g. 'You receive 100 CHF'). Ten different starting amounts were used in the experiment (from 10 CHF to 100 CHF with an increment of 10 CHF). To move on to the decision screen, the participants had to fixate for more than 2 seconds the received monetary amount, which was positioned in the centre of the screen. The decision in each trial consisted of choosing between a sure option and a gamble. The sure option was presented in the Gain frame trials as an amount of money retained from the starting amount (e.g. keep 70 CHF out of a total of 100 CHF) and in the Loss frame trials as the total amount of money lost from the starting amount (e.g. lose 30 CHF out of a total of 100 CHF). The gamble option was identical for both frames, and it always offered the chance to keep all of the starting amount of money with some probability. Seven different probabilities were used in the study, such that the probability of winning (or losing) in a given trial was either 30%, 40%, 50%, 60%, 70%, 80% or 90%. The expected outcomes – expected value (EV) - of sure and gamble options were always equivalent in each trial, and also mathematically equivalent between frames. However, the EV varied across trials from 3 to 90 CHF.

The sure and the gamble options on the screen were represented through rectangles, pie-charts and colours. Rectangles indicated the amount of money that the subject could either keep or lose from the starting amount, pie charts represented probabilities (e.g. the sure option always had a pie-chart completely filled with a colour), and colours were informative of whether the subject could keep or lose the amount of money from the starting amount indicated by the initially presented rectangles (e.g. the gamble always had one of the two 'keep' colours and never a 'lose' colour).

In the last phase of the experiment, our participants had to perform a certainty equivalence (CE) elicitation task in which they were shown all the 70 gambles that they encountered during the choice task. In each trial – for a total of 70 trials, a gamble and a list of sure amounts of money was shown to the participant (as an example see figure A.2). The subject had to make 10 decisions in each trial between the gamble displayed on the left hand-side of the screen and 10 sure amounts of money. The sure amounts of money were displayed in a decreasing order from a maximum value smaller but close to the amount of money that they could win if they choose the gamble to a minimum value close to zero. For instance, in the example in figure A.2, one the gamble can be described as follows. You first receive an endowment of 100 CHF and you have a probability of 70% of keeping all the 100 CHF

(30% of having nothing). In the first line, the participant had to decide between the gamble and a sure amount of 95 CHF. In the second line, she had to decide between the gamble and a sure amount of 85 CHF, and so on. The CE was calculated as the mean between the two amounts of money where the subject switches from choosing the sure amount to choosing the gamble. Thus, in the example in figure 2 if a participant chooses the sure amounts of money for every decision lines until 55 CHF, and she chooses the gamble from 45 CHF on, the CE for this gamble is 50 CHF.

We forced participants to only switch from choosing the sure amount to choosing the gamble and only once per trial – including the possibility of never switching, i.e. always choosing the gamble or always choosing the sure amount. The subjects were incentivized for the experiment with real money. They were informed that one of the two tasks – i.e. the choice task or the CE elicitation task – was randomly selected and implemented as a payment at the end of the experiment. If the choice task was selected the subject would receive the average amount of money of all the 150 realized choices that the subject made. In case the elicitation task was selected, the payment was given by the following procedure. One out of the 10 choices in each trial was randomly selected and implemented, then the payment was calculated as the average of all the 70 randomly selected choices. Furthermore, subjects were instructed to treat each decision separately and independently of the others.

All tasks were programmed in Matlab 2015b (Matworks), using the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007).

Eye-tracking

Before each decision trial, subjects were required to fixate the monetary endowment positioned at the center of the screen for 2 s before the options would appear, ensuring that subjects began every trial fixating on the same location.

Subjects' gaze was recorded at 500 Hz with an EyeLink-1000 (http://www.sr-research.com/) eye tracker. Choice trials with no gaze time on any option attribute were excluded from the analysis (17 trials, 0.004% of the pooled data from the 30 subjects).

Data analysis

We used the R package for statistical analysis of the behavioral results from the decision task (lme4 extension) and model estimations. The eye-tracking data were processed using the same procedures used in Krajbich et al., 2010 to analyze eye movements in a binary choice task.

Model fitting

We fitted the gain and loss condition separately with a hierarchical Bayesian method for the aDDM (equation A.1 and A.2). The code was implemented in R with extension package RJags for the hierachical estimation of the parameters, you can find the code on GitHub (`https://github.com/galombardi/method_HtSSM_aDDM`). In the hierachical Bayesian method, we make the standard assumption that individuals are members of a normally distributed population and assign a normal prior for each individual level parameter. Furthermore, before fitting the model we ran recovery fitting analysis to make sure the model was able to correctly fit the data. You can find the recovery procedure in the method paper (in preparation).

Then, we simulated choices and reaction time for loss and gain conditions with the mean of the individual posterior distributions for each parameter and each subjects. Figure SA4 shows the goodness of fit of the aDDM.

A.6 Figures and tables

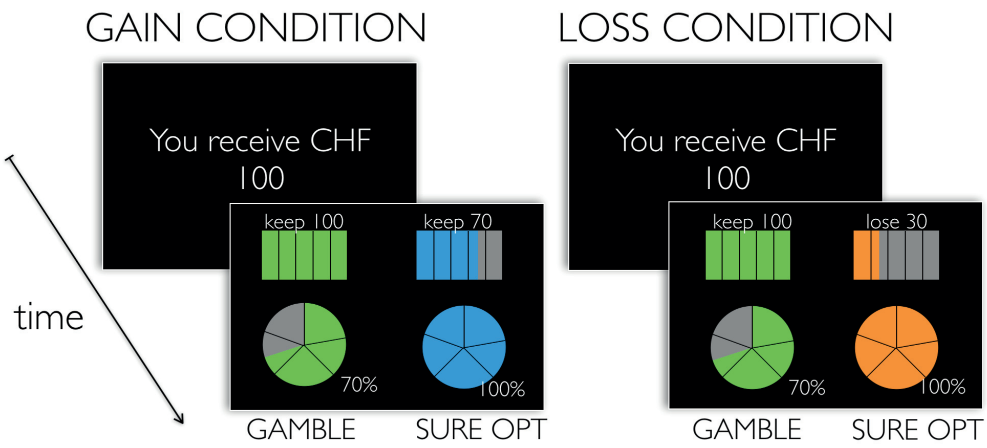


Figure A.1: Binary choice task. Subjects had to fixate on the first screen the initial amount of money they received in each trial for 2 s before seeing the decision screen that showed them the choice set. The decision screen presented the sure option (on the right-hand side of the screen in this example) and the gamble (on the left-hand side of the screen in this example). After a selection was made, a white box highlighted the chosen option for 1 s. In both the gain and the loss condition, the probability with which the subject could win or lose money was indicated by the colored part of the probability pie. In the gain condition, the amount of money subjects could win from an option was always indicated by the colored part of the rectangle. In the loss condition, the colored part of the rectangle for the sure option (in the above example this is the option that involves the fully orange-colored probability pie) indicates the amount of money from the initial endowment that the subjects lost for sure if they chose that condition. Notice that subjects were not shown the numbers and words written in the example above; they could only see shapes and colors. They learned, however, the meaning of the colours, rectangles and pie charts in two training sessions that enabled them to quickly understand the options.

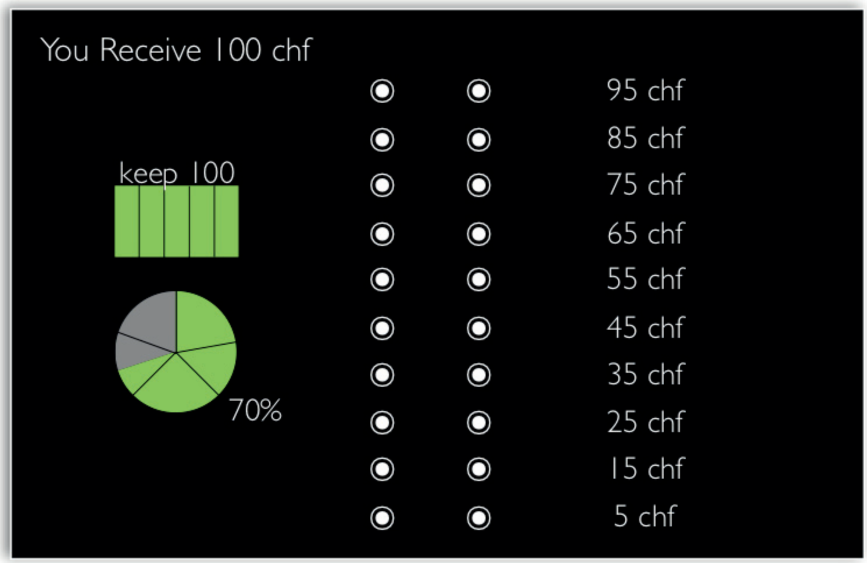


Figure A.2: Certainty equivalent elicitation task. Subjects were presented with all the possible gambles that they encounter in the choice task and asked to make a series of decisions for each of them as shown in the example screen. For each row they had to decide whether they prefer the gamble or a sure amount of money indicated on the right. For example, in the first row of the screen in the figure, they had to decide whether they prefer to keep 100 Swiss Francs that they received with 70% probability (or nothing) or if the prefer 95 Swiss Francs for sure; in the second row, they had to decide whether they wanted to keep 100 Swiss Francs with 70% probability (or nothing) or a sure amount of money of 85 Swiss Francs, and so on. They were also instructed not to make incoherent decisions, and to switch from choosing the sure amount of money to choosing the gamble not more than one time for each decision list.

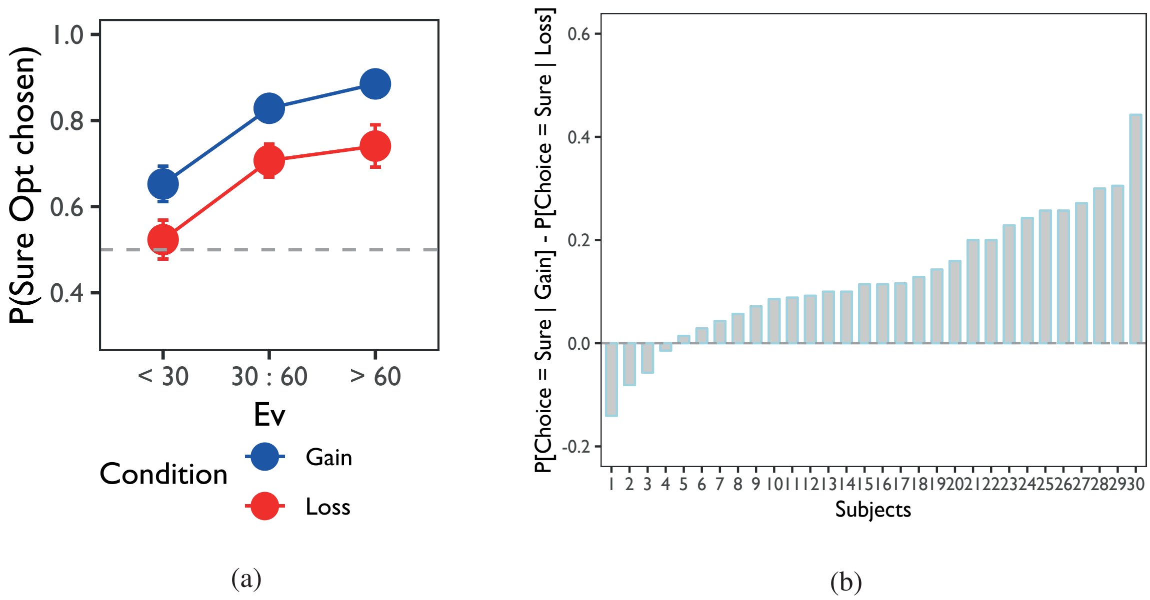


Figure A.3: Behavioral evidence for the framing effect. a) Probability of choosing the sure option as a function of expected value of the options for the gain and the loss condition. At every level of the options' expected value the sure option is chosen more frequently in the gain condition. The error bars indicate the standard errors of the mean. b) The behavioral influence of framing at the individual level. Each bar represents one individual and shows the difference between the probability of choosing the sure option in the gain condition and the loss condition. Values that are close to zero represent a minor framing effect.

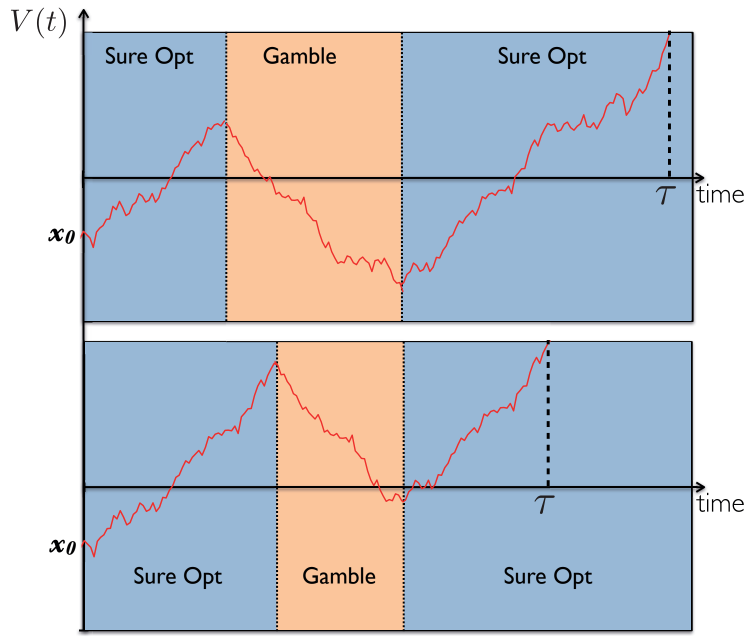


Figure A.4: Example of how a fixation time advantage for one option influences the reaction time. The plots show two examples of the relative decision value (RDV) for two options with the same values. The parameters of the model are kept constant across the two examples. The plots differs exclusively in the first fixation time duration, which is shorter in the upper plot than in the lower plot. In the figures, the colored part represents a fixation time duration, in blue for the sure option and in orange for the gamble. In both plots, the RDV crosses the upper barrier, meaning that the sure option has been chosen. Clearly, the decision time in the lower plot is smaller (i.e. faster choice) than the decision time in the upper plot, and this results exclusively from the fixation advantage towards the sure option in the lower plot.

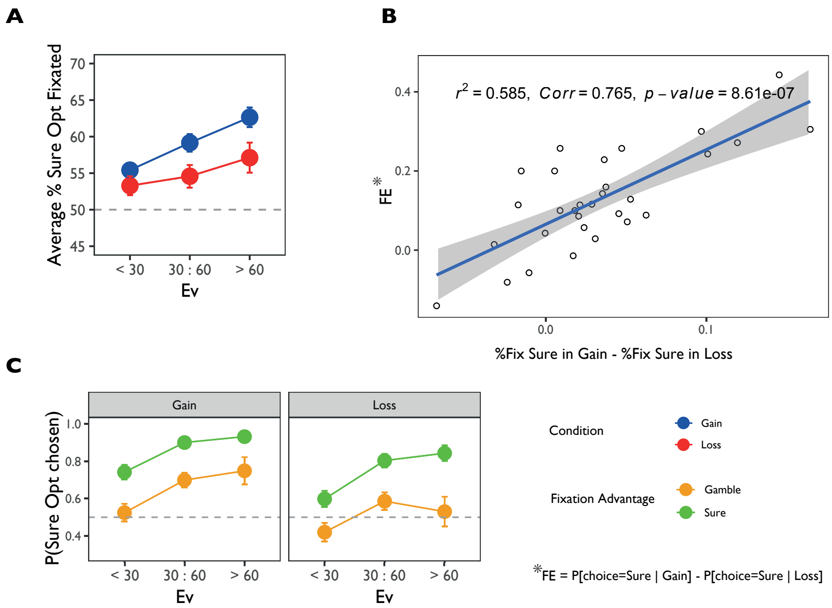
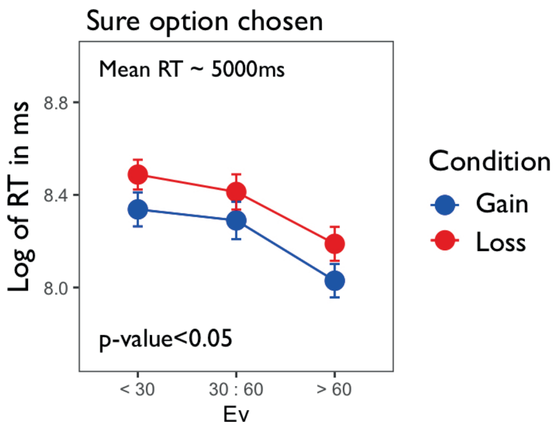
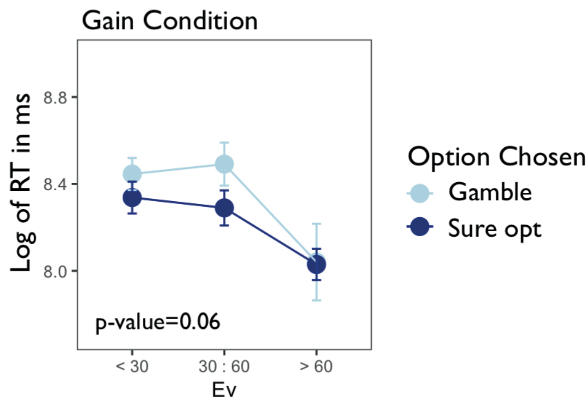


Figure A.5: The impact of framing on visual-attention and choices. In all figures, the error bars indicate the standard errors of the mean. (A) Average percentage of fixation time towards the sure option as a function of the options' expected value for gain and loss conditions. (B) Correlation between a measure of frame influence on choices (i.e. the difference between the probability of choosing the sure option in the gain condition and the loss condition) and the frame influence on fixations towards the option (i.e. the difference in percentage of fixation time towards the sure option between gain and loss). Each point represents a subject in the plot. (C) The probability of choosing the sure option separately for gain and loss conditions, as a function of expected value and for trials in which the sure option has a fixation advantage (green lines) and for trials in which the gamble has a fixation advantage (in orange).

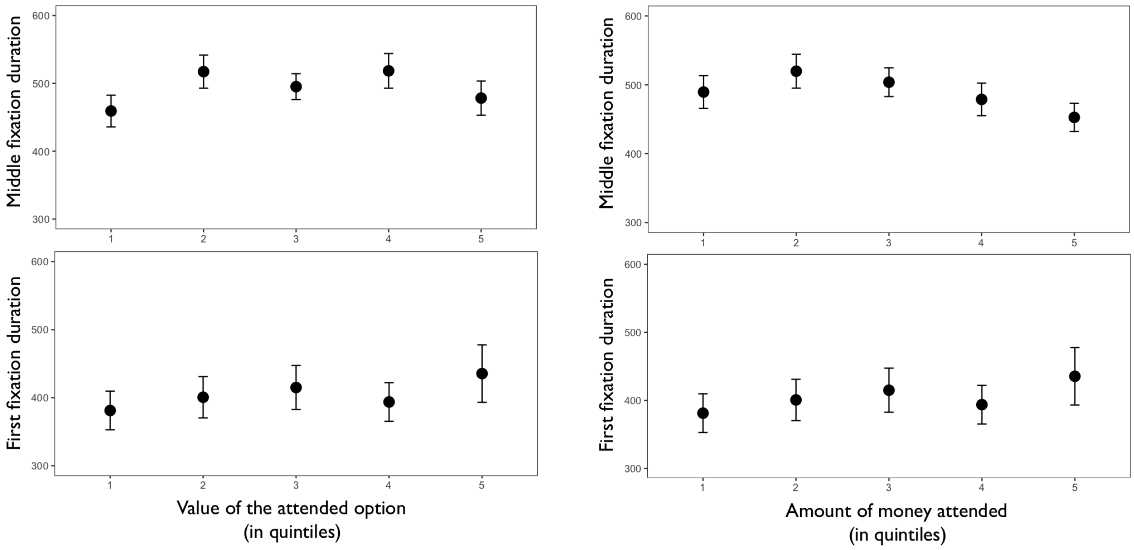


(a)



(b)

Figure A.6: The impact of framing on response times. In both figures, the error bars indicate the standard errors of the mean. (a) Average logarithm of the reaction time in milliseconds as a function of expected value Ev for gain (in blue) and loss (in red) conditions, only for trials where the sure option has been chosen. (b) Average logarithm of the reaction time in milliseconds as a function of expected value for trials where the gamble has been chosen (in light blue) and the sure option has been chosen (in dark blue).



(a) (b)

Figure A.7: Fixation duration as a function of values. (a) In the upper plot the average middle fixation duration as a function of the value of the option attended split in quintiles (value = CE if the attended option is the gamble and value = M_s if the attended option is the sure option); in the lower plot the average first fixation duration as a function of the value of the attended option split in quintiles (value = CE if the attended option is the gamble and value = M_s if the attended option is the sure option). (b) In the upper plot the average middle fixation duration as a function of the amount of money attended split in quintiles; in the lower plot the average first fixation duration as a function of the amount of money attended split in quintiles.

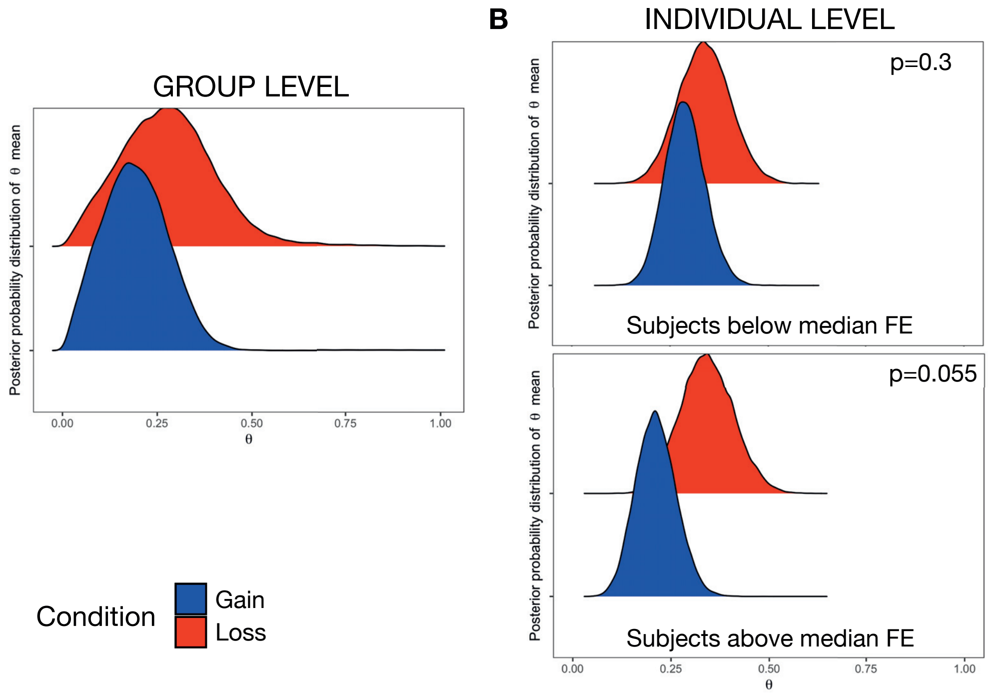


Figure A.8: Posterior distributions of the mean attentional discount parameter of the hierarchical Bayesian model fittings for the aDDM separately for gain (in blue) and loss (in red) conditions. (A) Posterior distributions of θ mean at the group level. (B) Posterior distributions of θ mean at the individual level of the HaDDM for two groups of subjects based on whether they showed a framing effect below or above the median (i.e. the framing effect is the difference in the probability of choosing the sure option between gain and loss conditions). The figure on top shows the posterior probability distributions of the individual level mean for the group of subjects (16) for which the framing effect is below median, and the figure on the bottom shows the posterior probability distributions of the individual level mean for the group of subjects (14) for which the framing effect is above median.

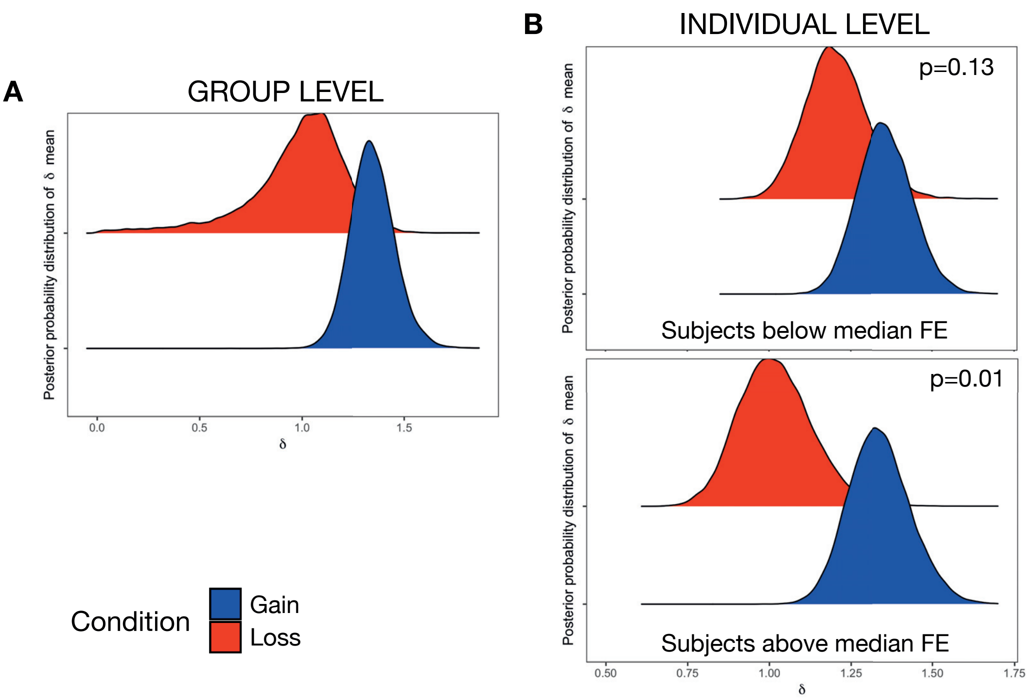


Figure A.9: Posterior distributions of the mean constant drift rate parameter of the hierarchical Bayesian model fittings for the aDDM separately for gain (in blue) and loss (in red) conditions. (A) Posterior distributions of δ mean at the group level. (B) Posterior distributions of δ mean at the individual level of the HaDDM for two groups of subjects based on whether they showed a framing effect below or above the median (i.e. the framing effect is the difference in the probability of choosing the sure option between gain and loss conditions). The figure on top shows the posterior probability distributions of the individual level mean for the group of subjects (16) for which the framing effect is below median, and the figure on the bottom shows the posterior probability distributions of the individual level mean for the group of subjects (14) for which the framing effect is above median.

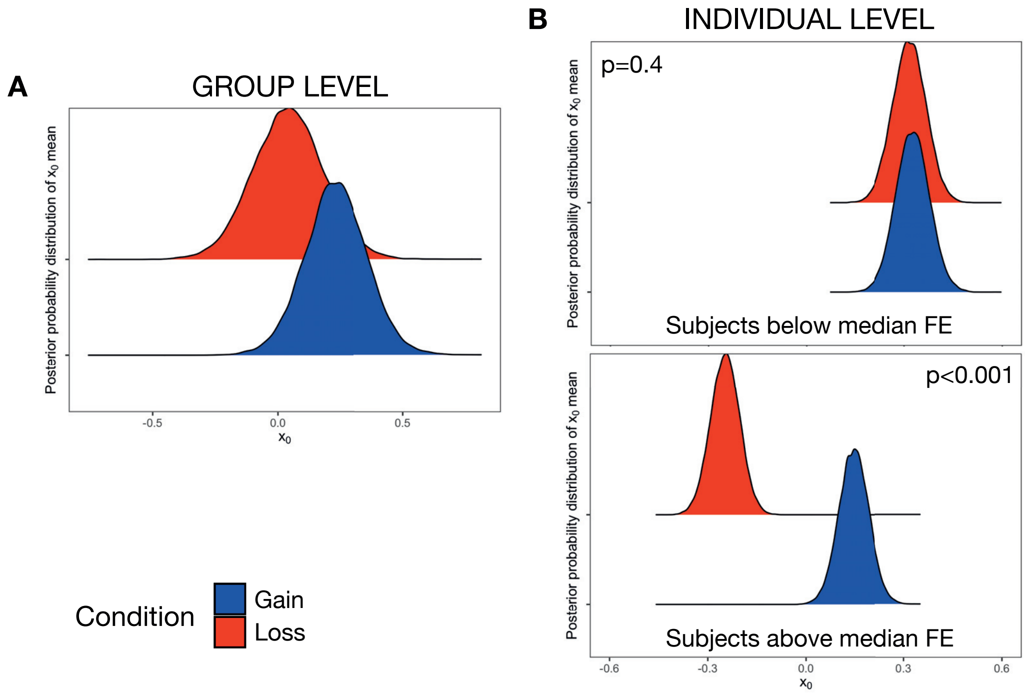


Figure A.10: Posterior distributions of the mean initial bias parameter of the hierarchical Bayesian model fittings for the aDDM separately for gain (in blue) and loss (in red) conditions. (A) Posterior distributions of x_0 mean at the group level. (B) Posterior distributions of x_0 mean at the individual level of the HaDDM for two groups of subjects based on whether they showed a framing effect below or above the median (i.e. the framing effect is the difference in the probability of choosing the sure option between gain and loss conditions). The figure on top shows the posterior probability distributions of the individual level mean for the group of subjects (16) for which the framing effect is below median, and the figure on the bottom shows the posterior probability distributions of the individual level mean for the group of subjects (14) for which the framing effect is above median.

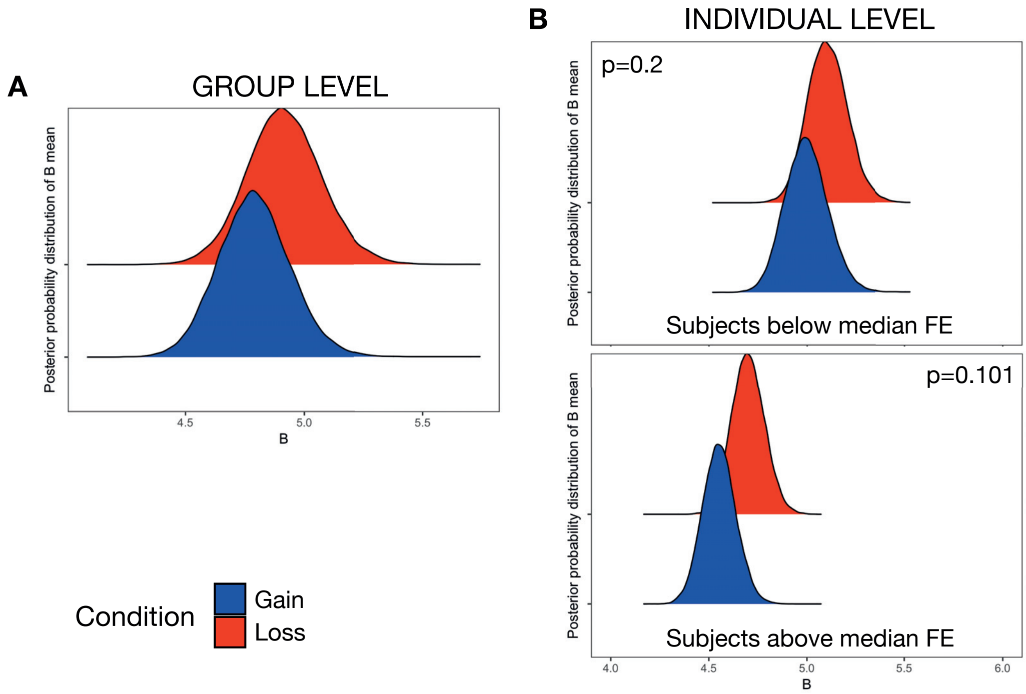


Figure A.11: Posterior distributions of the mean threshold parameter of the hierarchical Bayesian model fittings for the aDDM separately for gain (in blue) and loss (in red) conditions. (A) Posterior distributions of B mean at the group level. (B) Posterior distributions of B mean at the individual level of the HaDDM for two groups of subjects based on whether they showed a framing effect below or above the median (i.e. the framing effect is the difference in the probability of choosing the sure option between gain and loss conditions). The figure on top shows the posterior probability distributions of the individual level mean for the group of subjects (16) for which the framing effect is below median, and the figure on the bottom shows the posterior probability distributions of the individual level mean for the group of subjects (14) for which the framing effect is above median.

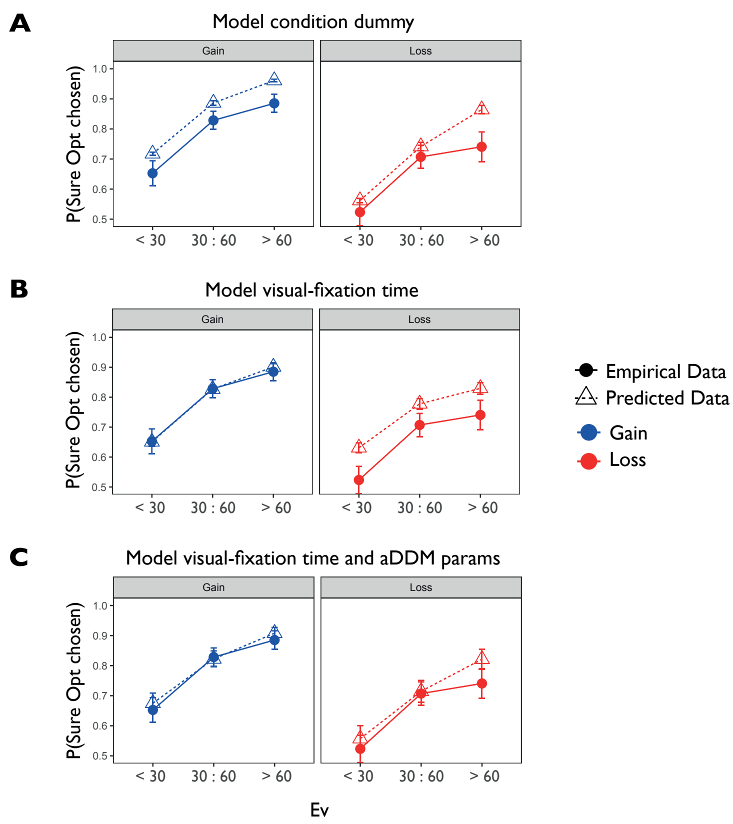


Figure A.12: Predicted and empirical probability of choosing the sure option for three different logistic mixed-effect regression models. (A) The upper-side model is the simple model in figure 7, in which the regressors are the following: a dummy variable for loss condition (baseline gain condition), value difference between sure option and gamble (M_s - CE), expected value (Ev = M_s), and the interactions between value difference and expected value with loss condition (Loss Condition X (M_s - CE) and Loss Condition X Ev, respectively). BIC = 3905.5, AIC = 4249.9. (B) The model in the middle is the model that takes into account only visual-fixations to explain the framing effect. The regressors for this model are the following: proportion of fixation time to the sure option, value difference between sure option and gamble (M_s - CE), expected value (Ev = M_s), and the interactions between value difference and expected value with the proportion of fixation time to the sure option. BIC = 4184.4, AIC = 3905.5. (C) The lower-side model takes into visual-fixation time and its interaction with the fitted parameters of the aDDM. The regressors for this model are the following: proportion of fixation time to the sure option, value difference between sure option and gamble (M_s - CE), expected value (Ev = M_s), the parameters of the aDDM (initial bias, attentional discount, drift rate and threshold), and the interactions between value difference, expected value and the parameter of the aDDM with the proportion of fixation time to the sure option. BIC = 4617.5, AIC = 3863.2. It is worth noticing that the last model is the one that better reproduce the data, without the need of a variable that indicates the frame condition (dummy condition for loss).

	Dependent variable:		
	Sure option chosen		
	(1)	(2)	(3)
Constant	1.719***	1.703***	1.297***
	(0.283)	(0.309)	(0.175)
Loss Cond	−0.930***	−0.894***	−0.182
	(0.174)	(0.184)	(0.187)
M_s	0.634***	0.564***	0.501**
	(0.163)	(0.168)	(0.172)
M_s - CE	0.623**	0.945***	0.785**
	(0.239)	(0.274)	(0.272)
Loss Cond × M_s	−0.263*	−0.138	−0.113
	(0.132)	(0.151)	(0.150)
Loss Cond × M_s - CE	−0.045	−0.281	−0.054
	(0.234)	(0.270)	(0.262)
% Fix sure opt		7.001***	6.659***
		(0.660)	(0.678)
Loss Cond × % Fix sure opt		−1.628*	−0.657
		(0.634)	(0.734)
M_s × % Fix sure opt		1.421**	1.086*
		(0.520)	(0.477)
M_s - CE × % Fix sure opt		1.261	1.525*
		(0.841)	(0.763)
Param x_0			1.100***
			(0.087)
Param θ			0.058
			(0.102)
Param δ			0.217*
			(0.104)
Param B			0.127
			(0.082)

	(1)	(2)	(3)
% Fix sure opt × x_0			−0.335
			(0.534)
% Fix sure opt × θ			-1.066^{*}
			(0.463)
% Fix sure opt × δ			0.582
			(0.531)
% Fix sure opt × B			0.591
			(0.447)
Log Likelihood	−2,097.969	−1,868.779	−1,804.472
Akaike Inf. Crit.	4,249.938	3,867.559	3,986.944
Bayesian Inf. Crit.	4,421.085	4,279.580	5,184.974
Pseudo-R^2	0.2041	0.2910	0.3154
Note:	. $p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$		

Table A.1: Three logistic mixed-effects regressions with random effects for subject-specific constants and slopes of the probability of choosing the sure option. Dependent variable equals 1 if the sure option is chosen and 0 otherwise. (1) The regressors for model 1 are the following: a dummy variable for loss condition (baseline gain condition), the value difference between sure option M_s and certainty equivalence of the gamble (M_s - CE), the value of the sure option (M_s = Ev), and the interactions between value difference and expected value with loss condition (Loss Condition X (M_s - CE) and Loss Condition X M_s, respectively). (2) In addition to the regressors of model 1, in model 2 the proportion of fixation time towards the sure option and all the possible interactions are investigated (% Fix Sure Opt, Loss Condition X % Fix Sure Opt, Loss Condition X (M_s - CE), and Loss Condition X Ev). (3) All the regressors from model 2 are in model 3 plus the parameters from the fitting of the HaDDM (one value per subject per condition), and their interaction with the proportion of fixation time to the sure option.

	Dependent variable:	
	% Fix sure opt	Fix Advantage sure opt
	Linear mixed-effects	*Logistic mixed-effects*
	(1)	(2)
Constant	0.578^{***}	0.657^{***}
	(0.009)	(0.086)
Loss Cond	-0.034^{**}	-0.252^{**}
	(0.009)	(0.091)
M_s	0.022^{***}	0.241^{***}
	(0.006)	(0.068)
M_s - CE	0.009	0.044
	(0.007)	(0.069)
Loss Cond × M_s	−0.011	-0.163^{*}
	(0.007)	(0.072)
Note:	. p<0.1; *p<0.05; **p<0.01; ***p<0.001	

Table A.2: Two mixed-effects regressions with random effects for subject-specific constants and slopes. (1) Linear mixed-effect model of the proportion of fixation time towards the sure option as a function of a dummy variable for the loss condition (baseline gain condition), expected value of the options (M_S = Ev), the value difference between sure option and gamble (M_S - CE), and the interaction of expected value and condition (Loss Condition $\times M_S$). (2) Logistic mixed-effect model of the probability that the sure option has a fixation advantage (= 0 if the gamble has a fixation time advantage, =1 if the sure option has a fixation advantage) as a function of the a dummy variable for the loss condition (baseline gain condition), expected value (M_S = Ev), value difference between sure option and gamble (M_S - CE), and the interaction of expected value and condition (Loss Condition $\times M_S$).

A.7 Supplementary Material

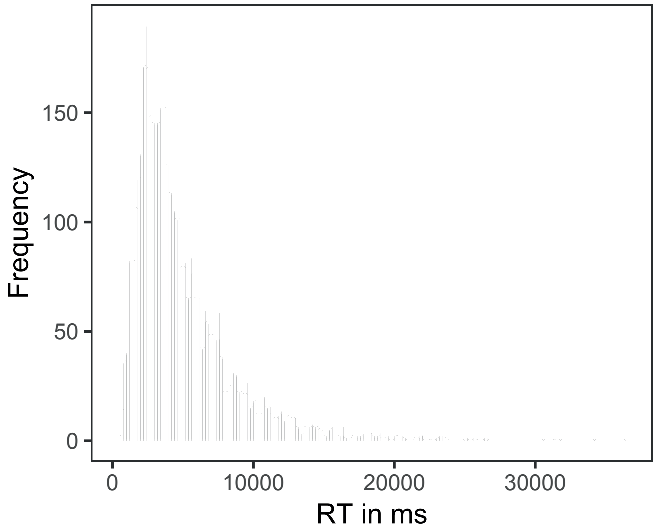


Figure SA1: Distribution of the reaction time.

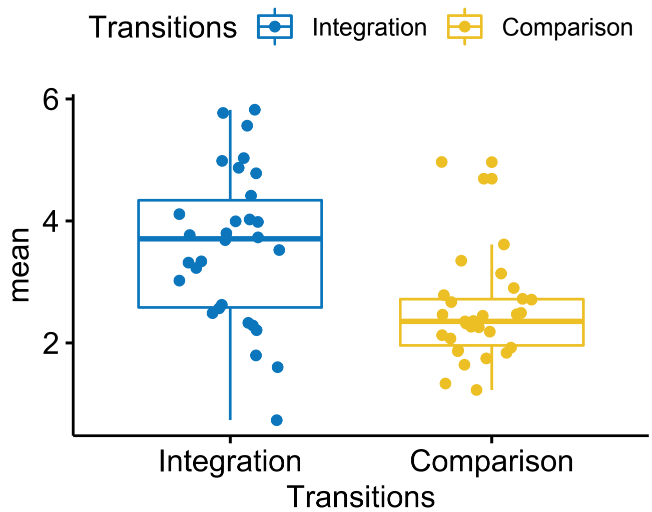
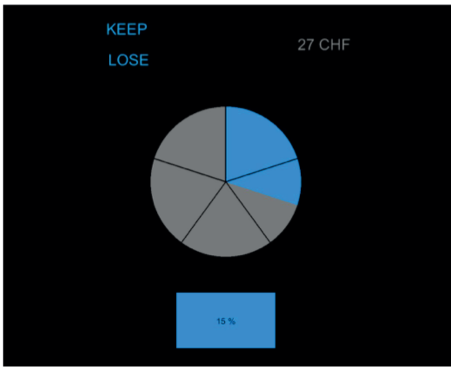
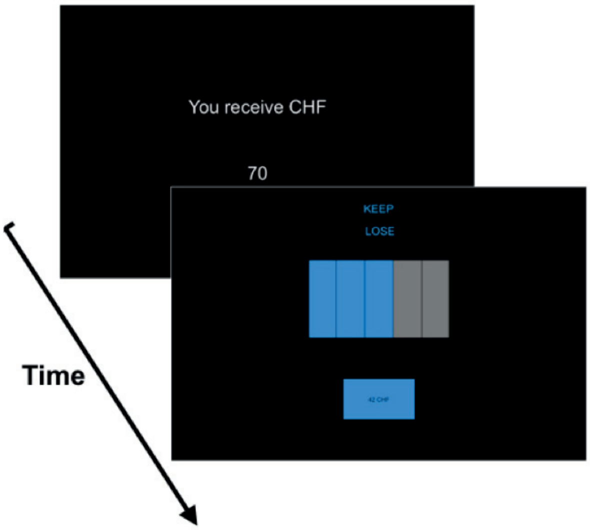


Figure SA2: Mean number of transitions for integration type (e.g. within an option) and comparison type (e.g. between options) of gaze patterns. Each point is a subject.



(a)



(b)

Figure SA3: Training phase. (a) Probability training phase. Subjects had to specify the probability represented by the colour-filled area of the pie-chart, and to indicate whether the colour had a 'keep' or 'lose' meaning. (b) Money training phase. Subjects had to specify the amount of money represented by the colour-filled area of the rectangle with respect to the received amount of money, and to indicate whether the colour had a 'keep' or 'lose' meaning.

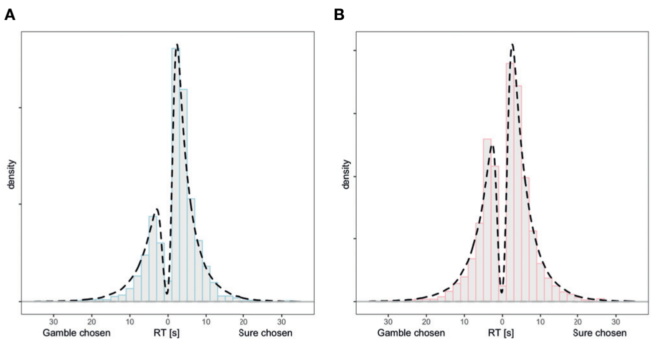


Figure SA4: Reaction time distribution of the empirical data (histograms) and simulated data (dashed lines) for (A) gain condition (in blue) and (B) loss condition (in red). In both plots, the reaction time is split by gamble chosen (from 0 to 30 seconds on the left-hand side) and sure option chosen (from 0 to 30 seconds on the right-hand side).

	Dependent variable:
	Logarithm of RT
Constant	8.270***
	(0.069)
Loss Cond	0.071*
	(0.030)
$\mid M_s$ - CE $\mid$	−0.064**
	(0.019)
Loss Cond × $\mid M_s$ - CE $\mid$	0.022
	(0.022)
Note:	. p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table SA1: A linear mixed-effects regression with random effects for subject-specific constants and slopes of the logarithm of the reaction time. The regressors for this model are the following: a dummy variable for loss condition (baseline gain condition), the absolute value of the value difference ($|M_s - CE|$), and the interactions between the loss condition dummy and the absolute value of the value difference (Loss Cond X $|M_s - CE|$)

	Dependent variable:	
	log of RT	
	(1)	(2)
Constant	8.284***	8.336***
	(0.074)	(0.067)
M_s	−0.079***	−0.081***
	(0.016)	(0.013)
Loss Cond	0.130**	
	(0.039)	
Loss Cond × M_s	−0.005	
	(0.023)	
Gamble Chosen		0.105
		(0.064)
Gamble Chosen × M_s		0.159***
		(0.025)
Note:	. p<0.1; *p<0.05; **p<0.01; ***p<0.001	

Table SA2: Two linear mixed-effects regressions with random effects for subject-specific constants and slopes of the logarithm of the reaction time. (1) In regression 1, we model only trials in which subjects choose the sure option. The regressors for this model are the following: a dummy variable for loss condition (baseline gain condition), expected value (M_s = Ev), and the interactions between the loss condition dummy and the expected value (Loss Cond X M_s). (2) In regression 2, all trials are considered. The regressors for model 2 are the following: a dummy variable for choosing the gamble (Gamble Chosen) for which the baseline is the sure option chosen, expected value (M_s = Ev), and the interactions between the Gamble Chosen dummy and the expected value (Gamble Chosen X M_s).

	Dependent variable:
	Count of Number of Transitions
Constant	0.875***
	(0.064)
Transitions Integration	0.289***
	(0.068)
Loss Cond	−0.034
	(0.037)
Ev	−0.109***
	(0.018)
Loss Cond × Transitions Integration	0.095**
	(0.030)
Transitions Integration × Ev	0.057**
	(0.021)
Loss Cond × Ev	0.059*
	(0.028)
Loss Cond × Transitions Integration × Ev	−0.009
	(0.029)
Note:	. p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table SA3: A poisson mixed-effects regression with random effects for subject-specific constants and slopes of the count of number of transitions' gazes from one attribute to the other per trials. The regressors of the model are the following: a dummy for the type of transition (baseline transition type comparison) which is 0 if the transition is between attributes of different options (money to money attribute or probability to probability attribute, i.e. transition comparison) and 1 if the transition is between attributes of the same option (money to probability or probability to money attributes, i.e. transition integration); a dummy for condition type (Loss Condition), the expected value of the option (Ev) and all possible interactions.

Subject	Condition	Mean # Trans Integration	Mean # Trans Comparison
1	Gain	**4.73**	**3.10**
2	Gain	**5.00**	**2.73**
3	Gain	**6.09**	**5.29**
4	Gain	**6.07**	**4.17**
5	Gain	**3.90**	**2.60**
6	Gain	**2.97**	**2.11**
7	Gain	**4.07**	**2.00**
8	Gain	**2.89**	**1.91**
9	Gain	**3.90**	**2.83**
10	Gain	**1.12**	**0.94**
11	Gain	**3.79**	**2.40**
12	Gain	1.96	1.93
13	Gain	**2.23**	**3.70**
14	Gain	0.81	2.37
15	Gain	**3.86**	**2.56**
16	Gain	**3.83**	**2.24**
17	Gain	**3.43**	**2.83**
18	Gain	**3.29**	**2.44**
19	Gain	**3.40**	**2.53**
20	Gain	**4.19**	**1.64**
21	Gain	**3.03**	**1.73**
22	Gain	**5.06**	**4.63**
23	Gain	**2.56**	**1.19**
24	Gain	2.99	3.10
25	Gain	**2.59**	**1.77**
26	Gain	**3.79**	**2.79**
27	Gain	1.82	1.93
28	Gain	**3.43**	**2.47**
29	Gain	**5.34**	**2.84**
30	Gain	**2.26**	**2.03**
1	Loss	**5.33**	**3.17**
2	Loss	**4.56**	**1.99**
3	Loss	**5.46**	**4.10**
4	Loss	**5.57**	**3.06**
5	Loss	**4.14**	**2.32**
6	Loss	**3.66**	**2.14**
7	Loss	**4.16**	**2.37**
8	Loss	**2.34**	**1.57**
9	Loss	**3.70**	**1.81**
10	Loss	**2.09**	**1.72**
11	Loss	**4.19**	**2.13**

Subject	Condition	Mean # Trans Integration	Mean # Trans Comparison
12	Loss	1.63	1.91
13	Loss	**2.19**	**3.00**
14	Loss	0.66	2.14
15	Loss	**4.13**	**2.39**
16	Loss	**5.00**	**2.74**
17	Loss	**3.61**	**2.59**
18	Loss	**3.39**	**2.26**
19	Loss	**4.06**	**2.91**
20	Loss	**5.56**	**2.03**
21	Loss	**3.43**	**1.56**
22	Loss	4.91	5.30
23	Loss	**2.41**	**1.27**
24	Loss	**3.06**	**2.70**
25	Loss	**2.55**	**1.96**
26	Loss	**3.59**	**2.56**
27	Loss	**2.83**	**2.22**
28	Loss	**4.12**	**2.42**
29	Loss	**5.79**	**2.73**
30	Loss	**2.31**	**1.73**

Table SA4: Participants mean number of transitions integration type (e.g. within an option) and comparison type (e.g. between options) of gaze patterns, for gain and loss conditions. In bold all the subjects for which the mean number of transition integration type is higher than the comparison type.

	Dependent variable:				
	Sure option chosen				
	(1)	(2)	(3)	(4)	(5)
Constant	1.297***	1.384***	1.623***	1.333***	1.357***
	(0.175)	(0.115)	(0.248)	(0.157)	(0.183)
Loss Cond	−0.182	−0.346*	−0.560**	−0.157	−0.321 .
	(0.187)	(0.145)	(0.207)	(0.156)	(0.181)
M_s	0.501**	0.575***	0.581**	0.510**	0.525**
	(0.172)	(0.159)	(0.180)	(0.163)	(0.169)
M_s - CE	0.785**	0.587**	0.611*	0.817**	0.846**
	(0.272)	(0.204)	(0.267)	(0.265)	(0.279)
Loss Cond × M_s	−0.113	−0.250*	−0.232	−0.126	−0.118
	(0.150)	(0.125)	(0.144)	(0.145)	(0.155)
Loss Cond × M_s - CE	−0.054	0.076	0.050	−0.053	−0.106
	(0.262)	(0.201)	(0.225)	(0.245)	(0.278)
% Fix sure opt	6.659***		6.249***	6.840***	6.727***
	(0.678)		(0.645)	(0.709)	(0.643)
Loss Cond × % Fix sure opt	−0.657		−0.734	−1.258 .	−0.829
	(0.734)		(0.720)	(0.752)	(0.683)
M_s × % Fix sure opt	1.086*		0.915*	1.195*	1.145*
	(0.477)		(0.458)	(0.495)	(0.480)
M_s - CE × % Fix sure opt	1.525*		1.236 .	1.578*	1.525 .
	(0.763)		(0.641)	(0.793)	(0.795)
Param x_0	1.100***	1.017***		1.094***	1.178***
	(0.087)	(0.091)		(0.090)	(0.085)
Param θ	0.058	0.045	0.191		0.002
	(0.102)	(0.086)	(0.144)		(0.119)
Param δ	0.217*	0.175*	0.420*	0.212*	
	(0.104)	(0.088)	(0.167)	(0.100)	
Param B	0.127	0.157*	0.309 .	0.177 .	0.083
	(0.082)	(0.062)	(0.183)	(0.097)	(0.083)

	(1)	(2)	(3)	(4)	(5)
% Fix sure opt × x_0	-0.335			-0.218	-0.574
	(0.534)			(0.550)	(0.511)
% Fix sure opt × θ	-1.066^{*}		-1.100^{*}		-1.104^{*}
	(0.463)		(0.439)		(0.491)
% Fix sure opt × δ	0.582		1.005 .	0.798	
	(0.531)		(0.521)	(0.604)	
% Fix sure opt × B	0.591		0.518	0.362	0.322
	(0.447)		(0.372)	(0.687)	(0.453)
Log Likelihood	$-1{,}804$	$-2{,}045$	$-1{,}835$	$-1{,}809$	$-1{,}813$
Akaike Inf. Crit.	3,987	4,219	3,974	3,921	3,929
Bayesian Inf. Crit.	5,185	4,631	4,938	4,885	4,893
Pseudo-R^2	0.32	0.22	0.30	0.31	0.31
Note:		. $p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$			

Table SA5: Five logistic mixed-effects regressions with random effects for subject-specific constants and slopes of the probability of choosing the sure option. Dependent variable equals 1 if the sure option is chosen and 0 otherwise. (1) Full Model. The regressors for model 1 are the following: a dummy variable for loss condition (baseline gain condition), value difference between sure option M_s and certainty equivalence of the gamble (M_s - CE), value of the sure option (M_s = Ev), the interactions between value difference and expected value with loss condition (Loss Condition X (M_s - CE) and Loss Condition X M_s, respectively), the proportion of fixation time towards the sure option and all the possible interactions are investigated (% Fix Sure Opt, Loss Condition X % Fix Sure Opt, Loss Condition X (M_s - CE), and Loss Condition X Ev), and the parameters from the fitting of the HaDDM (one value per subject per condition), and their interaction with the proportion of fixation time to the sure option. (2) Model without attention. The regressors of model 2 are the same as model 1 expect the proportion of fixation time towards the sure option and all the possible interactions. (3) Model without the initial bias. The regressors of model 3 are the same as model 1 expect the initial bias and its interaction with the proportion of fixation time. (4) Model without the attentional discounting factor. The regressors of model 4 are the same as model 1 expect the attentional discounting parameter and its interaction with the proportion of fixation time. (5) Model without the constant drift parameter. The regressors of model 5 are the same as model 1 expect the constant drift parameter and its interaction with the proportion of fixation time.

	Dependent variable:		
	Sure option chosen		
	(1)	(2)	(3)
Constant	1.719***	1.219***	1.118***
	(0.283)	(0.252)	(0.102)
Loss Cond	−0.930***		
	(0.174)		
M_s	0.634***	0.383***	0.415**
	(0.163)	(0.116)	(0.134)
M_s - CE	0.623**	0.759***	0.652***
	(0.239)	(0.184)	(0.163)
Loss Cond × M_s	−0.263*		
	(0.132)		
Loss Cond × M_s - CE	−0.045		
	(0.234)		
% Fix sure opt		6.559***	6.069***
		(0.548)	(0.501)
M_s × % Fix sure opt		0.846	0.898 .
		(0.544)	(0.485)
M_s - CE × % Fix sure opt		1.918*	1.392*
		(0.923)	(0.687)
Param x_0			1.057***
			(0.087)
Param θ			−0.029
			(0.088)
Param δ			0.278**
			(0.093)
Param B			0.138
			(0.091)

	(1)	(2)	(3)
% Fix sure opt × x_0			-0.162
			(0.464)
% Fix sure opt × θ			-1.082^{**}
			(0.410)
% Fix sure opt × δ			0.897 .
			(0.459)
% Fix sure opt × B			0.478
			(0.432)
Log Likelihood	$-2{,}098$	$-1{,}909$	$-1{,}813$
Akaike Inf. Crit.	4,249.938	3,905.509	3,863
Bayesian Inf. Crit.	4,421	4,184	4,617
Pseudo-R^2	0.20	0.28	0.31
Note:	. $p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$		

Table SA6: Three logistic mixed-effects regressions with random effects for subject-specific constants and slopes of the probability of choosing the sure option. Dependent variable equals 1 if the sure option is chosen and 0 otherwise. (1) Model condition dummy. The regressors for model 1 are the following: a dummy variable for loss condition (baseline gain condition), value difference between sure option M_s and certainty equivalence of the gamble (M_s - CE), value of the sure option (M_s = Ev), the interactions between value difference and expected value with loss condition (Loss Condition X (M_s - CE) and Loss Condition X M_s, respectively (2) Model visual-fixation time. Model 2 is similar to model 1 but the loss condition dummy is replaced by the proportion of fixation time. (3) Model without the initial bias. The regressors of model 3 are the same as model 2 plus the parameters from the fitting of the HaDDM (one value per subject per condition), and their interaction with the proportion of fixation time to the sure option.

	Dependent variable:	
	Log of Fixation Duration	
	Middle fixation	First fixation
Constant	5.807^{***}	5.604^{***}
	(0.034)	(0.061)
Loss Cond	0.046^{*}	0.070^{***}
	(0.020)	(0.024)
Option Value	0.022	0.028
	(0.014)	(0.024)
Loss Cond × Option Value	−0.027 .	−0.017
	(0.016)	(0.011)
Note:	. p<0.1; *p<0.05; **p<0.01; ***p<0.001	

Table SA7: Two linear mixed-effects regressions with random effects for subject-specific constants and slopes of the logarithm of fixation time duration. (1) Middle fixation duration. The regressors for model are the following: a dummy variable for loss condition (baseline gain condition), the value of the attended option (i.e. if gamble attended Option Value = CE, if sure option attended Option Value = M_s) and their interactions. (2) Initial fixation duration. The regressor for the model are the same as the model for middle fixation duration.

Appendix B: Study 2

Is the decoy effect an attention-driven phenomenon?

Gaia Lombardi, Todd Hare, Ernst Fehr

University of Zürich

Department of Economics

Zürich Center for Neuroeconomics (ZNE)

B.1 Abstract

The decoy effect is a well documented example of a preference reversal in which individuals seem to change their subjective valuation of two options (A vs B) when a third, irrelevant, alternative (C) is introduced in the choice-set. However, the mechanisms underlying the decoy effect are still imperfectly understood. Why should an irrelevant change in the choice-set affect decision-making?

Here, we examine the hypothesis that introducing a third alternative in the choice-set causes changes in the allocation of visual-attention to the options, and in turns visual-attention changes give rise to changes in the decision process. We investigate and replicate the attraction effect in a lottery task, i.e.,, individuals change their risk attitude as a function of whether the decoy relates to option A or B. Although the decoy is basically never chosen it has a considerable effect on the relative allocation of attention to the two relevant options. If the decoy similar is to option A it induces frequent comparisons between the decoy and this option and thus strongly increases the relative attention to option A. We then ask how and by which mechanisms these shifts in visual-attention allocation could influence choices. Based on an sequential sampling process approach (race model and aDDM), we test different modeling hypothesis that could explain the decoy effect through a reallocation of visual-attention without changing the valuation of the options in the choice set.

B.2 Introduction

In classical economics, the standard theory of rational choice assumes that preferences between alternatives do not depend on the context or on the presence of other options. This principle assumes that the decision maker has a complete preference order of all options and she will always choose the higher ranked option in this order. However, experimental evidence shows that this assumption is violated in many choice contexts (Tversky 1972, Huber and Puto 1982, Tversky and Simonson 1993; Shafir 1993). One notable example of this violation is the decoy effect which occurs when the introduction in the choice set of a third irrelevant option changes the preference relation between the other two options (Tversky 1972; Huber, Payne, and Puto 1982; Simonson, 1989; Slaughter et al. 1999; Soltani et al. 2012).

Although the decoy effect has been known in economics since 40 years and also more recently has been widely investigated in psychology and in neuroscience, which tried to unveil the brain and psychological processes during this type of decisions, there is still scarce consensus on the driving mechanisms for such effect, and the conditions under which decision makers exhibit different forms of this violation are a matter of current debate.

Many computational modelling attempts to explain the various types of decoy effects have been attempted (Kahneman and Tversky, 1984; Busemeyer and Townsend, 1993; Hotaling et al., 2010; Roe et al., 2001; Usher and McClelland, 2004; Bhatia, 2013; Noguchi and Stewart 2018; Gluth et al. 2018), the majority of which are based on sequential sampling models. For the past two decades sequential sampling models have become dominant in value-based decision making, mainly because of their feature of predicting not only choices but also reaction time distribution (Ratcliff, R. et al. 2016; Forstmann et al. 2016; Hanks and Summerfield 2017). The fundamental principle of these models is that during the decision process the decision maker accumulates evidence for each option until it reaches a certain threshold at which point a decision is made.

Also, more recently in neuroscience and psychology, the idea that attention plays a role in the decoy effect and more in general in decision making has gained popularity (Krajbich and Rangel, 2011; Tsesos, Chater and Usher 2012; Gluth, Spektor and Rieskamp 2018; Noguchi and Stewart 2018) and it has been incorporated in some sequential sampling models (Krajbich and Rangel, 2011; Roe et al., 2001; Usher and McClelland, 2004; Bhatia, 2013; Noguchi and Stewart 2018; Gluth et al. 2018). However, in most of these computational theories attention plays a secondary role in explaining context-dependent effect such the decoy. Addittionally, in very few studies that investigate the decoy effect (Nouguchi and Stewart 2014; Gluth et al. 2018) attention has been fully recorded.

Based on growing literature on the importance of attention in value-based and perceptual decision making (Krajbich et al. 2010; Krajbich and Rangel 2011; Krajbich and Rangel 2012; Towal et al. 2013), we hypothesize that attention could play a more important role in explaining the decoy effect than previously believed. The study aims to seek evidence to support the assumption of a key role of attention in the decoy effect as inner mechanims of the decision process. Thus, we combined a behavioral paradigm, eye-tracking and computational techniques to examine the role of visual-fixations as a proxy for attention in decoy-dependent changes of preferences during risky decision problems.

We developed a lottery task in which subjects performed a series of trinary decisions between risky options while their gaze movements were recorded. One of the options was always the irrelevant decoy, i.e., always dominated in money and probability by one of the other two options, which we refer to here as the target. This task was designed to capture the attraction effect, i.e., a decoy effect in which the decoy is never chosen and it attracts more choices for the target option, in a within-subject design fashion which allowed us to record changes in visual-attention allocation across two different decoy conditions.

Behaviorally, we replicated the attraction effect, i.e., the target option was chosen more often than the competitor option. Concerning the visual-attention allocation, we first found a bias in fixation time towards the target option which was strong and significantly correlated to the strengths of the decoy effect at the individual level. Additionally, we recorded an effect on choice frequency of looking time to the decoy. Surprisingly, the probability of choosing the target was significantly influenced by the time participants spent attending the decoy. We investigated this result further and find evidence suggesting that the decoy alters attention allocation within the choice set making the target option the center of attention. From our result, it seems reasonably clear that visual-attention plays a major role in the decoy effect. However, what is not entirely clear is the mechanisms with which attention acts to influence choices.

Because our experiment was not designed to rule out and isolate the determinants of the probability of attending the options in the choice set nor to define the proper relation between attention, options' values and decoys, we tested the predictions of other models in our data. Thus, we reviewed the models in the literature that include attention in their framework and examine different predictions of how attention is allocated among the options in the choice set, and how it could influences the decoy effect. We find that only very few predictions of the examined models seem to be in line with our eye-gaze data, suggesting a more complicated and still unsolved role of attention in the decoy effect.

In summary, our results suggest a fundamental role of the decoy in attracting and shaping attention allocation which was not previously recorded. The decoy causally manipulates the attention allocated by the decision maker on options in the choice set and this shifts in attention are strictly related to choice frequency in very specific ways that have not been previously accounted for.

B.3 Results

Experimental Design

The main goal of this study is to investigate the attentional processes that could play a role in the decoy effect. In particular, we chose to examine the attraction effect which seems to be one of the most robust preference reversals among the decoy effects, and simultaneously record decision makers' eye movements.

Participants in our experiment performed a series of trinary choices where they had to decide between three lotteries in each trial: option A, option B and a decoy. Each option, if chosen, would always give an amount of money with a certain probability p or nothing with probability 1-p. During the all task, option A was the option with high money and low probability attributes, and B was the low money and high probability option, whereas the decoy could be either of type A in half of the trials or of type B in the second half of the trials (see figure B.1a). Irrespectively of its type, the decoy was always a dominated option. Namely, in half of the trials both attributes of the decoy were smaller than the attributes of option A (condition 1, see Figure B.1a for an example), and in the other half of the trials they were smaller than the attributes of option B (condition 2 in figure B.1a). Whenever the decoy is of type A we refer to option A as the target option and option B as the competitor. Vice-versa, when the decoy is of type B, option B is the target option and A the competitor. Extensive evidence in economics and psychology (e.g., Huber, Payne, and Puto 1982; Shafir 1993; Soltani et al. 2012) highlight a causal role of this type of decoy in increasing the choice probability for the target option. Thus, in this study we refer to the decoy effect (DE) as the difference in probability between choosing the target and choosing the competitor.

It is worth mentioning that on the decision screen of the choice-task, participants in our experiment were not shown the numbers and symbols as we displayed in figure B.1a. They could only see the shapes and the colors representing the lottery options. This feature of our design allowed us to have longer fixation durations and to minimize the attentional component that might come from memory, because it is much easier to remember the number and letters than their graphic representation. However, we ensured that subjects knew the meaning of the shapes and colors with two extensive training phases before we started the main experiment (see Methods section for details)

Choice Behavior

As expected, we find a decoy effect in our experiment. More specifically, we observed the predicted attraction effect, i.e., the target was chosen significantly more often than the competitor (figure B.2a). The decoy effect seems also to be very robust across participants (figure B.2b). The data also show that there is an overall tendency to choose option B – the high money and low probability option – more often than option A. This is most likely due to the fact that the expected value of option B (EV B) was always higher than the expected value of option A (EV A). We designed the task in such a way to achieve approximate indifference in choosing between option A and option B in the absence of any decoy. The result shows that in fact we did not succeed in producing this indifference level between A and B, however, we obtained a significant decoy effect on choice probability overall.

Furthermore, we show that contrary to what was stated in Frederick et al. 2014, the decoy effect can be found without numeric representation of the options as in our experiment.

Eye-movements results

Because we measure subjects' gazes with the eye-tracker, we can use patterns of visual-fixations to provide a deeper mechanistic understanding of the decoy effect in the decision makers decision process. Specifically, we want to study how visual-attention is influenced by different decoys and how this in turns relates to choices and reaction time. There is a growing body of evidence on the role of attention in decision making suggesting a prominent role of attention on choice frequency. In particular, visually guided attention has been shown to bias the decision process in favour of the alternative that is attended longer (Ashby et al., 2016; Cavanagh et al., 2014; Konovalov and Krajbich, 2016; Kovach et al., 2014; Krajbich and Rangel, 2011; Krajbich et al., 2010; Lim et al., 2011; Schonberg et al., 2014; Shimojo et al., 2003; Stewart et al., 2016; Towal et al., 2013; Vaidya and Fellows, 2015). Considering that we observed in our data a bias in choice frequency towards the target, based on this previous evidence we can expect a bias in attention towards this option. Thus, we investigate whether the decoy not only leads to changes in choice frequency between A and B, but whether it also drives changes to the allocation of attention between among the two options.

As expected, we find a significant bias in looking time towards option A or B depending on whether the decoy is of type A or B, respectively. Namely, we register a decoy effect at the fixation level, i.e., on average a positive fixation time difference between target and competitor. Figure B.3 and table SB2 in the supplementary material show the proportion of fixation time as a function of the decoy type.

Furthermore, we are interested in understanding how the resulting bias in fixation time to the target relates to choice frequency. We find that the probability of choosing the target can be predicted by the fixation time advantage to the target (beta = 1.6 and p-value < 2e-16, see table SB1 model 2 and 3 in the supplementary material), and that one standard deviation increase in the proportion of fixation time to option B produces a three times larger effect on choice probability of option B than one standard deviation increase of value difference (EV B – EV A, beta = 0.49, p-value < 5e-06) and the decoy type dummy (beta = 0.3, p-value = 0.052).

In addition to this strong effect of the fixation time bias on choice frequency, table SB1 in the supplementary material show that when controlling for proportion of fixation time to option B and proportion of fixation time to the decoy, the direct effect of the decoy type almost vanishes completely, suggesting that attention is able to account for the variability in the choice probability. Further evidence for this hypothesis can be find in figure B.4a. We see here that the decoy effect can be reversed for trials in which the fixation time bias is reversed: in trials in which the competitor is attended longer than the target we find on average a higher choice probability for the competitor than the target (see figure B.4a).

The previous evidence suggests a fundamental role of visual-attention on the decoy effect and also provides support for the view that attention might play more than a secondary role on this type of preference reversals. However, if fixation time has an impact on choice frequency, this should also hold at the level of individual trials, regardless of the decoy type. Namely, in choice trials in which an option is systematically advantaged in terms of fixation time, this option should have a higher probability of being chosen relative to the choice trials in which an other option has an advantage in fixation time. Indeed, when splitting trials depending on which option in the choice set is the longest attended option, we find that independently of the decoy type trials in which option B has a fixation time advantage have a higher choice probability in favor of option B compared to the trials in which A is the longest attended option (figure B.4c). Moreover, this holds even when splitting trials by which option is the last attended option, see results in figure B.4b.

After observing such a high impact of fixation time advantages on choice probability, and more specifically on the probability of choosing the target, we examine the questions of how and why the decoy makes the target the longest attended option in our task. One possible hypothesis is that the decision maker is trying to first resolve the dominance relation between target and decoy, and this attempt induces more fixations towards these two options. When investigating eye-gaze transitions between options, we find that the number of transitions between target and decoy is significantly higher than those between competitor and decoy (figure B.5a and table SB3 in the supplementary material). Also, the probability of transitioning from decoy to target conditional to looking at the decoy is much larger than the probability of transitioning to the competitor (figure B.5b shows the conditional probability of eye-graze transitions). This suggests that the decoy makes the target the center of attention since most of eye-gaze transitions from and to the competitor or the decoy go via the target which is indeed the longest attended option in the choice set.

Moreover, when analyzing the eye-gaze dynamics during trial time, we find that the time the decoy is attended significantly decreases when approaching a decision (figure B.6). This provides further evidence to the hypothesis that decision makers in our experiment seem to resolve the dominance relation first and then focus more on the target and competitor options to make the choice.

According to some theory and empirical evidence on the relation between visual-attention and subjective values (Krajbich et al. 2010; Krajbich and Rangel 2011), the eye-gaze dynamics amplifies the evidence for the attended options during the decision process. Specifically, the more a decision maker looks at an option the more likely she is to choose that option, and this is in some cases independently of the value of the options. In our experiment however, the decoy is attended on average 25% of the decision time, but it is almost never chosen. Thus, what happens to the choice frequency when the decoy is attended? Does the simply decoy drive more fixations to the target or does attending the decoy have an impact on the choice frequency between target and competitor?

We investigate the probability of choosing the target as a function of the proportion of fixation time to the decoy and find that this has an impact on the probability of choosing the target. Namely, the decoy effect on choices significantly increases with the proportion of fixation time to the decoy (figure B.7a and table SB1 model 2 and 3 in the supplementary material), even when controlling for how long the target is attended (table SB1 model 2 and 3 in the supplementary material). We also find that both looking at the decoy and target are related to the decoy effect at the individual level. Once the decoy is attended long *enough*, the decoy effect on choices, i.e., the difference between the probability of choosing the target and the probability of choosing the competitor, is highly correlated with the decoy effect on fixations, i.e., the bias in fixation time to the target with respect to the competitor. Figure B.7b shows the correlation between decoy effect on choices and decoy effect on fixations for subjects split by median proportion of fixation time to the decoy. We find that the proportion of fixation time to the decoy amplifies the impact of fixation time difference between target and competitor on choice probabilities. This effect is also evident when running a mixed-effect regression model on the decoy effect on choices as a function of proportion of fixation time to the decoy and proportion of fixation time difference between target and competitor, i.e., decoy effect on fixations (see table SB4 in the supplementary material).

Modelling attention

After analysing eye-movements and transition probabilities it seems reasonably clear that attention plays an important role in the decoy effect. Thus, we turn to the literature and investigate the models which incorporate attention in their framework. We ask whether these models are in line with our data, and which are the most prominent and plausible mechanisms through which attention operates to affect choices in the decoy effect (for a detailed review of the models see Models summary in the supplementary material, Turner et al. 2017 or Busemeyer et al. 2019).

The most prominent theories which make assumptions on attention are the following: the multialternative decision field theory (MDFT; Busemeyer and Townsend, 1993; Hotaling et al., 2010; Roe et al., 2001) model, the multiattribute leaky competing accumulator (MLCA; Usher and McClelland, 2004) model, the associations and accumulation model (AAM; Bhatia, 2013), the multialternative decision by sampling model (MDbS, Noguchi and Stewart 2018), the mutual inhibition value-based attention capture (MIVAC, Gluth et al. 2018), and the attentional drift diffusion model (aDDM, Krajbich and Rangel 2011).

The MDFT and MLCA models have similar predictions on how attention should be allocated among options' attributes in the choice set. Even though, attention allocation does not have a prominent role in explaining the decoy effect in these models, they predict that the accumulation process is performed at the attribute level and this seems to be important to explain other decoy effects like the similarity effect. The decision maker switches from accumulating evidence for one attribute to accumulating evidence to the other attribute in a race model fashion.

Thus, if decisions are implemented following such mechanistic processes, we would expect the decision makers to compare the options at the attribute level. Namely, if during the decision the evidence is accumulated for each attribute separately, in each time step the decision maker has to sample one attribute at the time, e.g., the money or probability attribute, and look at that attribute across options to estimate the distance between them and integrate it in the accumulation process. However, we find that in our experimental data subjects significantly look more within than across options, thus performing more integration type eye-gaze transitions than comparison type transitions. In other words, participants in our task seem to sample information through eye-gaze in a way that suggests they integrate the values of the attribute for each option instead of separately comparing the values of the attributes among the options. Figure B.8 (see also table SB3 in the supplementary material for the regression model) shows the mean number of transitions types, i.e., integration and comparison transitions, for each subject in the experiment split by decoy type. Moreover, we notice not only that the eye-gaze transition type does not depend on the decoy type, but also the probability of attending one or the other attribute is not affected by the type of decoy present in the choice set (figure B.9). This evidence suggests that in our experiment it is unlikely that attention allocation could influence the choice frequency through a bias in fixation time at the attribute level.

On a similar line as the MDFT and MLCA, the AAM model makes assumptions on attention allocation at the attribute level. However, contrary to MDFT and MLCA this model provides detailed predictions about the probability of attending a specific attribute which depends on the value of the attended attribute and also on the values of the unattended ones. In particular, it predicts that the probability of attending attribute i has to be proportional to the overall value of all options in that attribute, and also negatively correlated to the sum of values of options in the other attributes. For example, the probability of attending the money attribute has to be proportional to the sum of money of all options and inversely proportional to the sum of probabilities of all options. When investigating in our data for this prediction, we find that the probability of attending the money attribute is significantly anti-correlated with the sum of all money and it is not significantly correlated with the sum of probabilities contrary to what the model would predict (see Fig B.10).

Another model which make testable predictions on attention allocation is the MDbS. This model predicts that similar options' attributes are more likely to be attended, and that the probability of attending the attribute of an option depends on the proximity to other options' attributes. In particular, the probability of attending attribute i of option A is proportional to the sum of the exponential of the proximity of A to all other options in the same attribute i, where the proximity to another option is defined as the negative relative distance, i.e., relative difference, from that option in attribute i (see Models summary in the supplementary material for more details). According to their definition, we calculated the exponential of the proximities for all options and attributes in our data set, i.e., option A, option B and decoy. We find that the probability of attending an option in an attribute depends as expected on the proximities to other options, but contrary to what the theory would predict, the sign of the relation between this probability and the proximities depends on the type of option we consider. For example, figure B.11 shows that the probability of attending the money attribute of option B depends positively on the proximity to the closest option and negatively on the proximity to the farthest option, and, also, the probability of attending the probability attribute of the decoy does not show a significant correlation with the proximity to any other options. Also, the model predicts that similar options should be attended more, thus resulting in a higher number of eye-gaze transitions between them. In our experiment, this translates in a higher number of eye-gaze transitions between target and decoy, which are the most similar options, compared to transitions between target and competitor. We find, however, a significantly higher number of eye-gaze transitions between target and competitor than transitions between target and decoy, suggesting that the two options that are attended the longest are target and competitor, and not target and decoy as the MDbS model would predict (see figure B.5 and B.6, and table SB3 in the supplementary material).

As described in details in the model summary section in the supplementary material, also the MIVAC model, which is a leaky sequential sampling model with inhibition, makes testable predictions about how attention allocation should influence choice frequency. First, the model assumes that the probability of attending an option is proportional to its value and, second, that attention gives a constant boost of evidence to the attended option, i.e., the subjective value of an option is increased by a constant value every time the decision maker is looking at it. In our data, however, we do not find the relation between the probability of attending an option and its value with or without controlling for the decoy type (figure B.12). Instead, the decoy seems to be the main source of attention manipulation in our data independently of the values of the options (figure B.12b). This model also makes a clear prediction on the interaction between the options' values and fixation time on choice frequency. The model assumes that during the evidence accumulation process attention impacts the evidence giving a constant boost to the value of the attended option. Thus, it clearly predicts that the higher the difference in values between the options the lower this impact of the attention boost on choice frequency should be. However, we do not see such negative interaction effect between options' value difference and fixation time on choice probability in our data (see table SB1 model 2 and 3 in the supplementary material).

Another sequential sampling model which makes predictions on the influence of attention allocation on choice frequency is the aDDM. Contrary to most of the models previously described, the aDDM does not make any assumption on the probability of attending an option; in the aDDM attention is an exogenous variable that drives the evidence accumulation process. In particular, the values of the non-attended options are discounted during the evidence accumulation process by a factor between 0 and 1, and this results in two main predictions on choice frequency. Firstly, the longer one option is attended the higher the probability of being chosen and a fixation bias, ceteris paribus, should result in a choice bias. Secondly, the model predicts an interaction effect between fixation time and the difference in the options' values on choice frequency: the higher the value difference between two options the lower the impact of fixation time on choice frequency. Also, the probability of choosing an option should depend on the interaction between the options' overall values and fixation time: the higher the sum of option values the higher the impact of fixation time on choice frequency. This follows from the assumption that the discount factor is multiplied with the value of the non-attended option. As shows in the previous section, we clearly find in our data that the probability of choosing an option increases with the fixation time to that option which is in accordance with the first prediction of the model. However, we do not see the predicted effects of the interaction between value difference and fixation time on choice frequency, and the interaction of options' values and fixation time (see table SB1 model 2 and 3 in the supplementary material). Also, the results displayed in figure B.7a, which shows the influence of fixation time to the decoy on the probability of choosing the target, cannot be explained with the standard form of the aDDM, suggesting that a more complex dynamics of values, visual-attention allocation and reaction time are involved in this type of decisions.

Overall our results suggest a more complicated and not fully understood effect of attention allocation on the choice process in the decoy effect. From our analysis it is rule out and isolate the determinants of the probability of attending the options in the choice set nor to define the proper relation between attention, options' values and decoys in this specific preference reversal effect. However, it seems clear from our results that the decoy its-self and not necessarily the values or the distance between options (see figure B.12b) is the main driver of visual-attention allocation, namely the different type of decoy induces the decision maker to allocate attention differently during the decision process, and this attention allocation bias seems to have a strong impact on choices.

B.4 Discussion

The decoy effect has been extensively investigated in economics and psychology for the last 40 years (Tversky 1972, Huber and Puto 1982, Tversky and Simonson 1993). However, the inner mechanisms driving this context-dependent effect are still unclear. Here, we hypothesized that attention has a more fundamental role in the decoy effect than previously recorded and we tested the hypothesis that the decoy effect changes the way attention is allocated among the options and that changes in the allocation of attention influence in turns choice frequency.

We recorded eye-gazes from individuals performing a trinary lottery choice-task which was designed to induce an attraction effect, i.e., a decoy effect in which the decoy is never chosen but attracts more choices for the target option. Indeed, we found that the different decoys drove different allocations of attention to the options in the choice set. In particular, the decoy not only attracted more choices to the target, but also increased the difference in fixation time between target and competitor. Furthermore, the decoy effect on choices seems to be highly correlated with the decoy effect on fixations, and we found that the decoy effect on choices could be completely reversed for trials in which the competitor was the longest attended option.

Adding more evidence to the hypothesis that visual-attention influences choice frequency in our task, we also found an effect of looking time on the probability of choosing an option that was independent of the decoy type. Namely, fixation time biases to option appeared to have an effect on choice probability for both types of decoy.

When investigating how and why the decoy made the target to be the longest attended option, we found evidence for the hypothesis that the decoy impacts the decision process by making the target the center of attention. First, the decoy forces more eye-gaze transitions from and to the target compared to the competitor. Second, the probability of transitioning from decoy to target conditional on looking at the decoy is much larger than the probability of transitioning to the competitor, and the probability of transitioning from the competitor to the target conditional on looking at the competitor is much larger than the probability of transitioning to the decoy.

One hypothesis for this effect is that decision makers have to resolve the dominance relation between target and decoy, and thus drives more eye-gaze transitions between target and decoy compared to target and competitor. Also, the looking time of the decoy in a trial significantly decreases in time, suggesting that once the dominance relation between target and decoy has been resolved the decision maker does not need to look at the decoy anymore.

Thus far all the evidence points to an important role of the decoy in shifting the attention allocation towards the target option making it the center of fixations and attracting more choices to it. This evidence reminds us of theories of the decision process, such as the aDDM, that attribute to attention allocation a causal role on choices. In other words, all the evidence seems to point to the prediction that decision-makers choose options they have looked at longer more often than would be predicted by their a priori options' values. However, we noticed that in fact the decoy option is attended on average 25% of the time. Thus, we asked whether this time only influences how long the target is going to be attended or whether this fixation time to the decoy has an impact on the decoy effect on choices. We found that the decoy effect on choices in our data is significantly increased by the proportion of fixation time to the decoy, even when controlling for looking time to the target. Hence, only looking at the decoy increases the probability that the target is chosen. In the logic of evidence accumulation processes, this means that when the subject looks at the decoy the evidence for the target is enhanced rather than reduced relative to the competitor. Also, the time spent looking at the target and the decoy are highly correlated at the individual level to the decoy effect on choices.

We then decided to examine whether there exists a sequential sampling model that incorporates attention in a way that can account for the type of behavioral and attentional patterns we found in our data. For example, one might speculate whether a different version of the aDDM in which different attentional discount factors for the different options are at work and could explain the observed effects. However, when testing the basic predictions of the aDDM model in our data we do not find evidence for such decision processes. Furthermore, even when testing the predictions of other sequential sampling models such as MDFT, MLCA, AAM and MDbS, we do not find evidence supporting any of these hypothesized decision processes in our data. Clearly, this data set is not sufficient to suggest a new model or any type of modification to the tested models. However, the evidence collected in this study undoubtedly points to a different type of mechanism relating visual-attention allocation and choice frequency in the decoy effect. It appears that the values of the options or the distances between options do not have such a clear and define role in attracting attention and thus in influencing the decoy effect. Whereas the presence of the decoy seems to be the main driver of attention allocation, and since it is the only aspect of the decision that changes across conditions, could be seen as a causal manipulation of visual-attention.

To conclude, three very interesting results have been recorded in this study about the decoy effect. 1) The magnitude of the decoy effect on choices changes accordingly to shifts in allocation of attentions to the target option and to the decoy its self. 2) The presence of the decoy influences eye-gaze transitions making the target the center of attention in the attraction effect. 3) None of the models in literature that incorporate attention in their framework seems to match the pattern of visual-fixations and choice frequency we find in our task. Thus, with this study and given the growing literature on the causal role of attention in choice (Shimojo et al. 2013; Armel et al. 2008; Tavares et al. 2017), we emphasize the need of investigating the mechanisms that, first, drive visual-attention and, second, impact choice frequency through changes in attention allocation during the decision making process.

B.5 Methods

Subjects

Forty-two healthy subjects participated in our experiment. Two of them were excluded from the experiment because the eye-tracker failed to record their eye-movements. Subjects received monetary compensation for their participation, were informed about all aspects of the experiment and gave written informed consent. The experiments conformed to the standards of the Declaration of Helsinki and the Human Subjects Committee of the University of Zürich approved the experimental protocol.

Task

The experimental design was divided into three parts: two training phases and a decision phase. During the first training phase, subjects were trained to learn how probabilities in the decision task were represented (figure SB2b) through several trials – the number of trials depended on the performance of each participant, from a minimum of 20 until the subject performed 10 corrected probability estimations in a row. In each trial, a computer displayed a colored pie-chart in the middle of the screen and box with a random percentage below it. A fraction of this pie-chart was filled in blue and the remaining part in gray. The size of the blue filled area in the pie-chart indicated a probability amount. In each training trial, subjects had to report in the box the exact probability represented by the colour-filled area.

In the second training phase, subjects learned to estimate monetary amounts (figure SB2a) through several trials – the number of trials depended on the performance of each participant, from a minimum of 50 until the subject performed 10 corrected money estimations in a row. In each trial, a computer displayed a colored rectangular shape in the middle of the screen and box with a random amount of money below it. As before, a fraction of this rectangle was blue and the remaining was filled in gray. The blue color indicated an amount of money and subjects had to report the exact amount of money in the box below the rectangle. A completely blue filled rectangle represented 10 Swiss Francs and a completely gray filled rectangle 0 Swiss Francs. The exact amount of money was equivalent to the proportion of blue filled area of the rectangle times 100 CHF.

After the training sessions, subjects performed the choice task. During this decision phase (figure B.1a), subjects had to make 136 decision trials between three lotteries: option A, option B and a decoy. Option A was always low money and high probability and option B always high money low probability, whereas the decoy was half of the trials of type option A or option B. However, the decoy was always dominated by option A or option B in both dimensions, money and probability. For example, the screen in figure 1 shows the following decision trial. Participants have to choose between the option B on the left-hand side, i.e., 95 CHF with 40% probability or nothing, option A on the right-hand side, i.e., 45 CHF with 80% probability or nothing, and the decoy in the middle down side of the screen, i.e., 40 CHF with 75% probability. In the example, the decoy is similar to option A and makes option A the target and option B the competitor. The task is a within subject design, thus participants faces the same option As and option Bs twice with different decoys, one similar to A and one similar to B. The amounts of money in the experiment ranged from 5 CHF to 100 CHF and the probability from 10% to 90%. The decoy was always at a fixed distance from target, i.e., always 5 CHF smaller in the money dimension and 10% smaller in the probability dimension.

The subjects were incentivized for the experiment with real money. They were informed that they would receive the average amount of money of all the 136 trials realized choices that they made. Furthermore, subjects were instructed to treat each decision separately and independently of the others.

All tasks were programmed in Matlab 2015b (Matworks), using the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007).

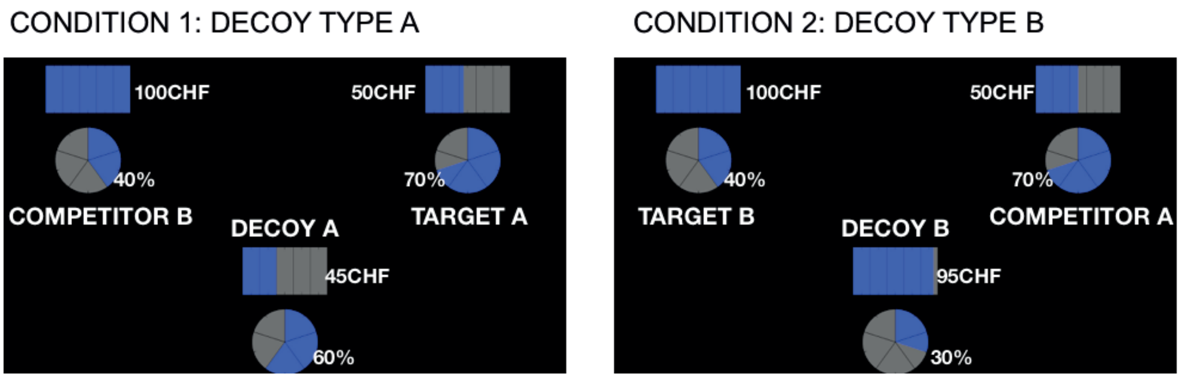
Eye-tracking

Before each decision trial, subjects were required to fixate a white cross positioned in the middle of the three options on a blank black screen for 2 s before the options would appear, ensuring that subjects began every trial fixating on the same location. Subjects' gaze was recorded at 500 Hz with an EyeLink-1000 (http://www.sr-research.com/) eye tracker. Choice trials with no gaze time on any option attribute were excluded from the analysis (2 trials, 0.035% of the pooled data from the 42 subjects).

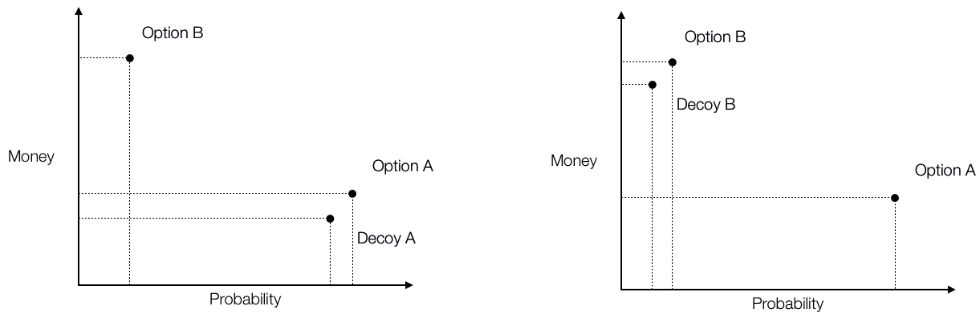
Data analysis

We used the R package for statistical analysis of the behavioral results from the decision task (lme4 extension) and model estimations. The eye-tracking data were processed using the same procedures used in Krajbich et al., 2010 to analyze eye movements in a binary choice task.

B.6 Figures and Tables



(a)



(b)

Figure B.1: a) Trinary choice task. Two examples of choice screens where subjects had to choose between a target, a competitor and a decoy options. In condition 1, the decoy was of type A, i.e., low money high probability, whereas in condition 2 it was of type B, i.e., high money low probability. In both conditions, the decoy was dominated either by option A or option B in both attributes. You can see the relation between decoys and options in b) where the options in the two conditions are displayed in money and probability dimensions. Notice, in a) numbers, percentages and words are for illustration purposes, the participants did not see them during the real experiment.

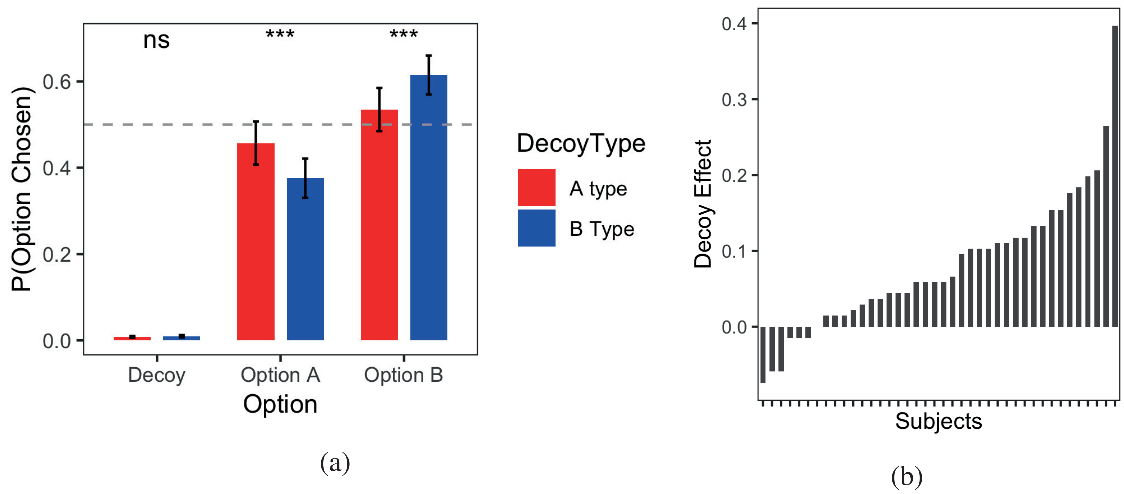
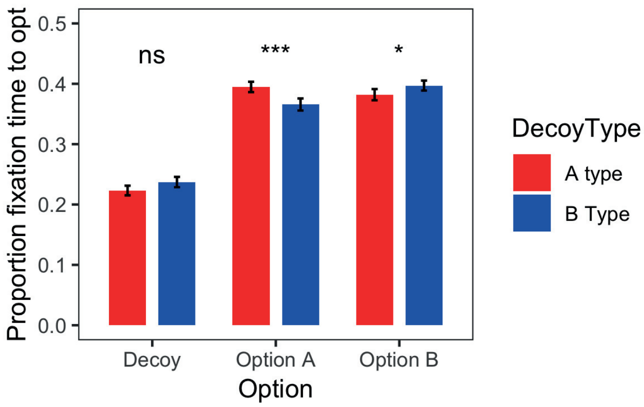
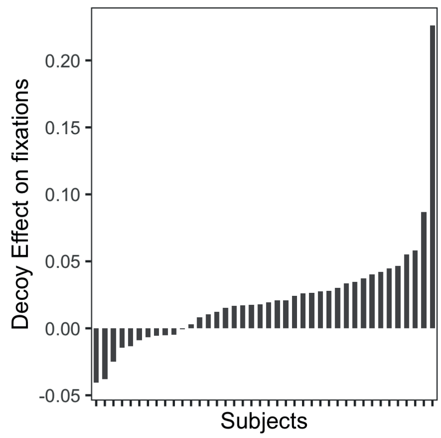


Figure B.2: a) Bar plot of the probability of choosing each option in the experiment split by condition – i.e., decoy type A and decoy type B. Note that the probability of choosing option A is significantly higher (see table SB1 in the supplementary material for the logistic mixed-effect regression model) when the decoy is of type A compared to when the decoy is type B, and viceversa, the probability of choosing option B is higher for trials of decoy B than trails of decoy A. Error bars indicate the standard error of the mean. b) The decoy effect displayed for each subject in the experiment. The decoy effect is defined as the difference between the probability of choosing the target and the probability of choosing the competitor. The attraction effect is when the DE>0, which is true for almost our subjects in the experiment.

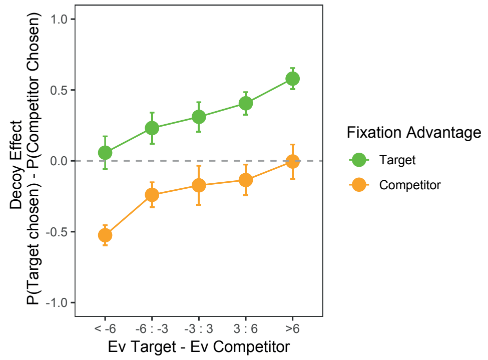


(a)

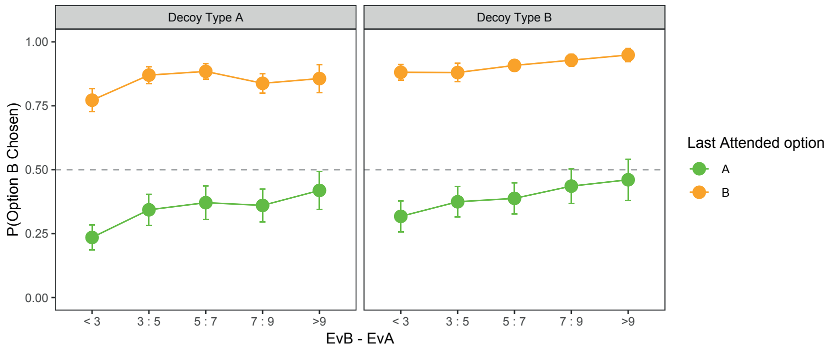


(b)

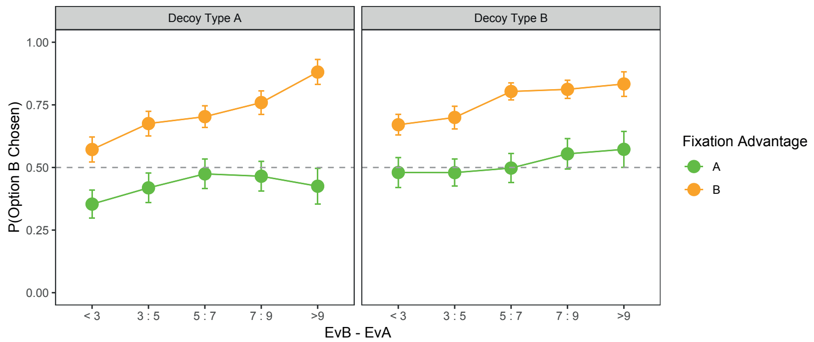
Figure B.3: a) Bar plot of the proportion of fixation time to each option in the experiment split by condition – i.e., decoy type A and decoy type B. Note that the proportion of fixation time to option A is significantly higher (see table SB3 in the supplementary material for the mixed-effect regression model) when the decoy is of type A compared to when the decoy is type B, and vice-versa, the proportion of fixation time to option B is higher for trials of decoy B than trails of decoy A. Error bars indicate the standard error of the mean. b) The decoy effect on fixations displayed for each subject in the experiment. The decoy effect on fixations is defined as the mean difference between proportion of fixation time to target and proportion of fixation time to competitor. See table SB2 for the regression models.



(a)

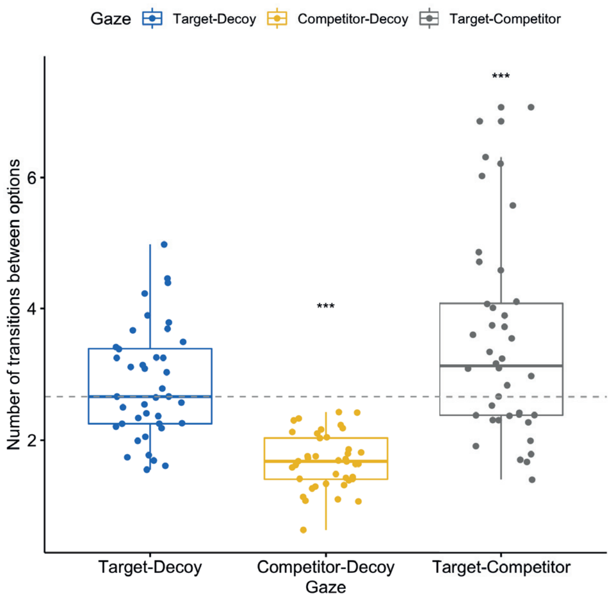


(b)

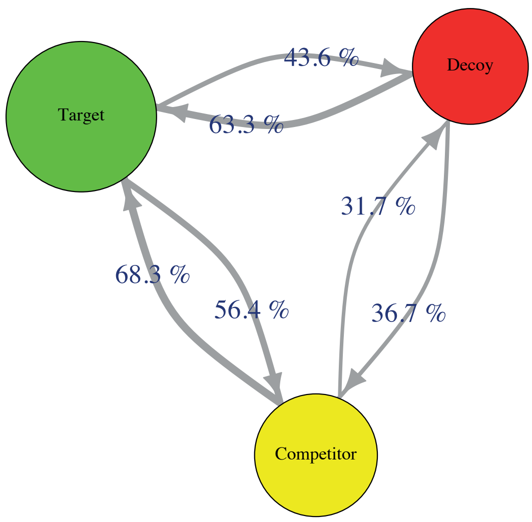


(c)

Figure B.4: a) Plot of the decoy effect on choices, i.e., the difference in probability between choosing the target and choosing the competitor, as a function of the EV difference between target and competitor, split by trials in which the target has a fixation time advantage (in green) and trials in which the competitor has a fixation time advantage (in orange). b) Plot of the mean probability of choosing option B as a function of the EV difference between option B and option A, split by trials in which the option A is the last attended option during the decision process (in green) and trials in which option B is the last attended option during the decision process (in orange). c) Plot of the mean probability of choosing option B as a function of the EV difference between option B and option A, split by trials in which the option A has a fixation time advantage (in green) and trials in which option B has a fixation time advantage (in orange).

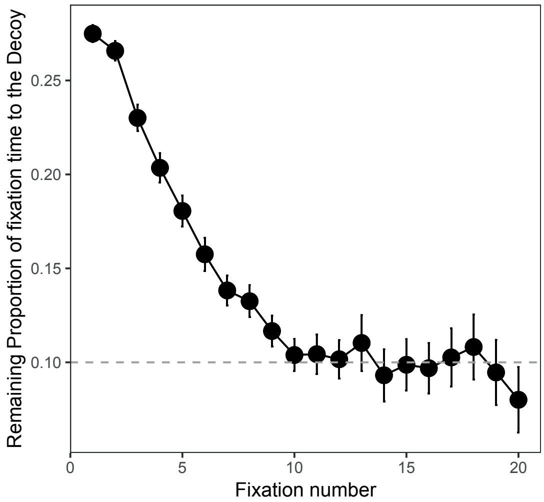


(a)

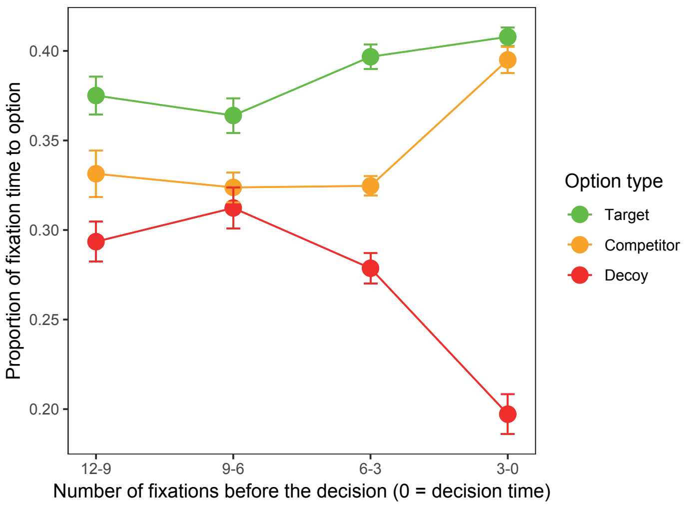


(b)

Figure B.5: a) Box-plot of the mean proportion of eye-gaze transitions between options. In blue the proportion of transitions between target and decoy, in yellow between competitor and decoy and in grey between target and competitor. b) Graph of the probability of eye-gaze transitions conditional on the fixating each option. Notice that the size of the circle for each option (in green the target, in yellow the competitor and in red the decoy) is scaled with the overall mean proportion of fixation time to each option.

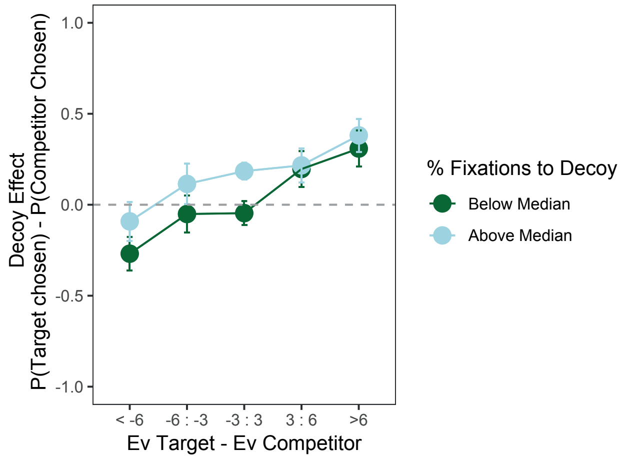


(a)

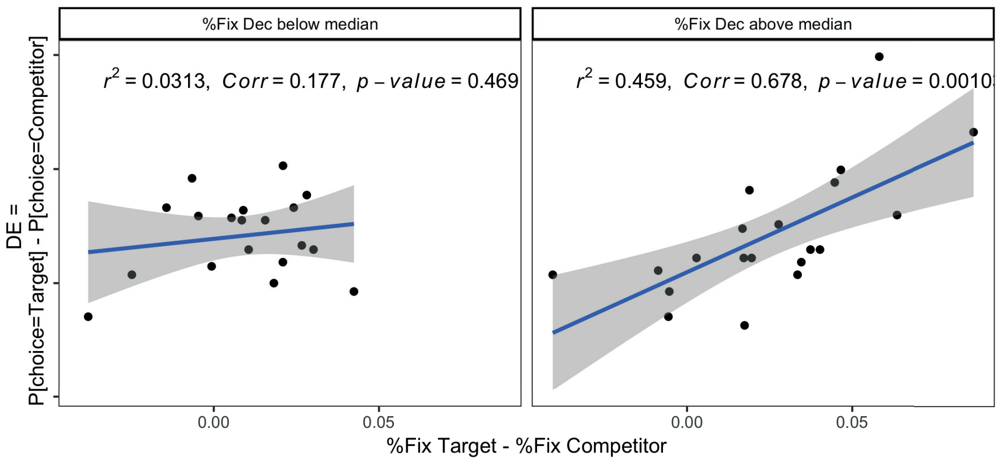


(b)

Figure B.6: a) Plot of the mean proportion of fixation time to the decoy before at a fixation number before the trial ends, i.e., in 0 the plot displays the mean proportion of fixation time to the decoy in the whole trial, and for instance, in fixation number 10 it displays the remaining mean proportion of fixation time to the decoy after 10 fixations. b) Plot of the mean proportion of fixation time to each option (target in green, competitor in yellow and decoy in red) in windows of fixation numbers before the end of a trial. For example, in 3 – 0 the plot displays the mean proportion of fixation time for each option in the last 3 fixations before the end of a trial.



(a)



(b)

Figure B.7: a) Plot of the decoy effect on choices as a function of the EV difference between target and competitor, split by trials in which the Decoy is attended below median percentage of the decision time (in green) and trials in which the Decoy is attended above median percentage of the decision time (in light blue). b) Scatterplot of the decoy effect on choices as a function of the difference in proportion of fixation time between target and competitor, split by trials in which the Decoy is attended below median percentage of the decision time (plot on the left-hand side) and trials in which the Decoy is attended above median percentage of the decision time (plot on the right-hand side). Note each point is a subject.

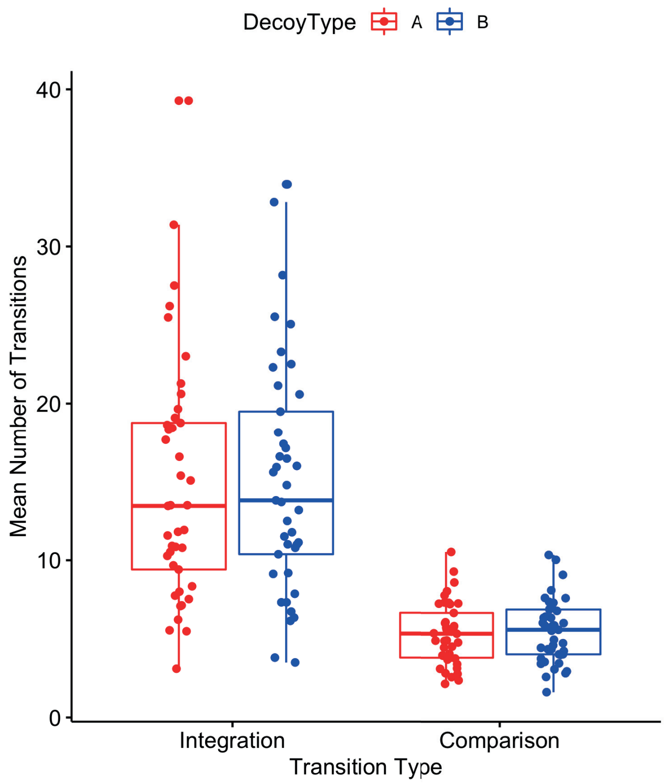


Figure B.8: Scatterplot of the mean number of gaze transitions within options (i.e., integration type of gaze) and across options (i.e., comparison type of gaze), splitting trials by decoy type. Each point is a subject and the plot displays median, quantiles and outliers.

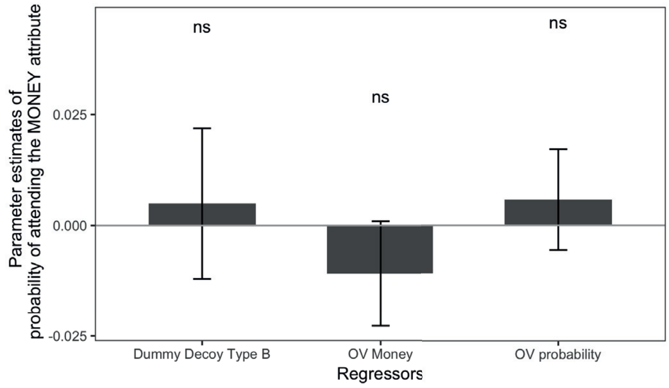


Figure B.9: Mixed-effect logistic regression model with random effects for subject-specific and trial-specific constants and slopes of the probability of attending the money attribute as a function of the a dummy variable for the decoy type (baseline Decoy Type = A), the sum of all money in each trial (OV money = Money A + Money B + Money Decoy) and the sum of probabilities in each trial (OV probability = Probability A + Probability B + Probability Decoy). Note, the interactions are not displayed in the plots but have been included in the regressions.

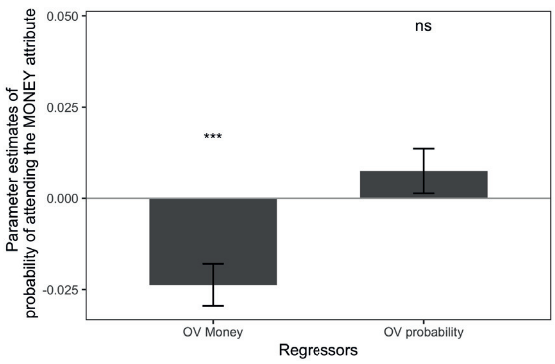
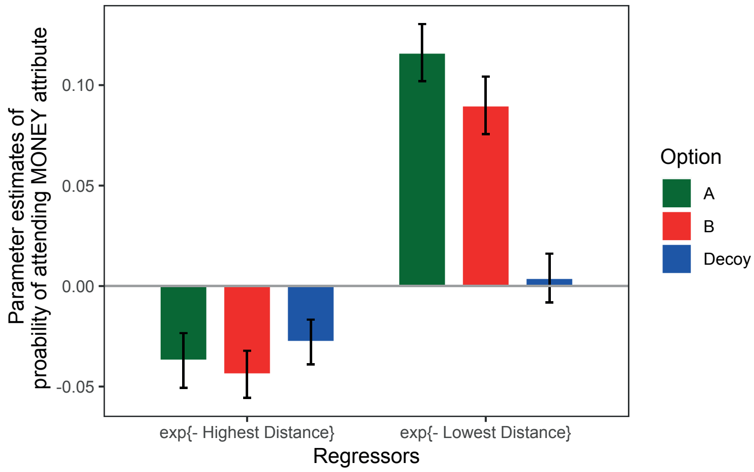
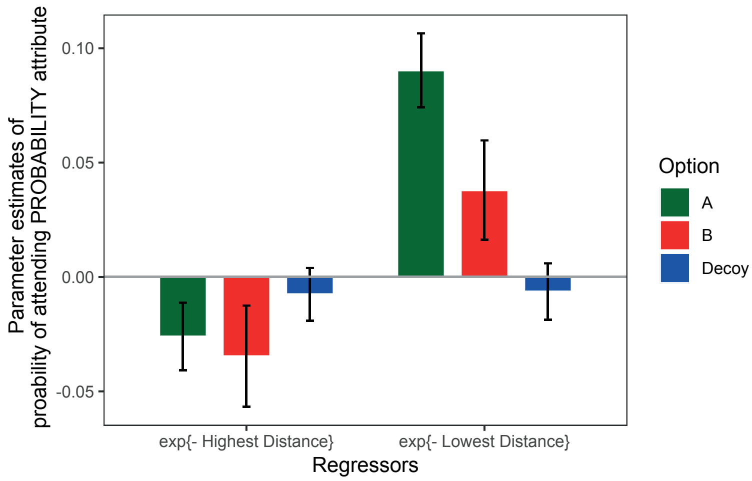


Figure B.10: Mixed-effect logistic regression model with random effects for subject-specific and trial-specific constants and slopes of the probability of attending the money attribute as a function of the overall values in money and probabilities– i.e., the sum of all the money values and the sum of all probability values (OV money = Money A + Money B + Money Decoy and OV probability = Probability A + Probability B + Probability Decoy, respectively). Note, the interactions are not displayed in the plots but have been included in the regressions.

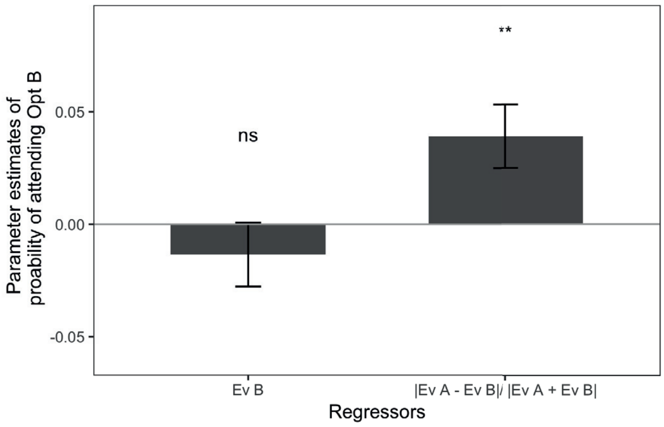


(a)

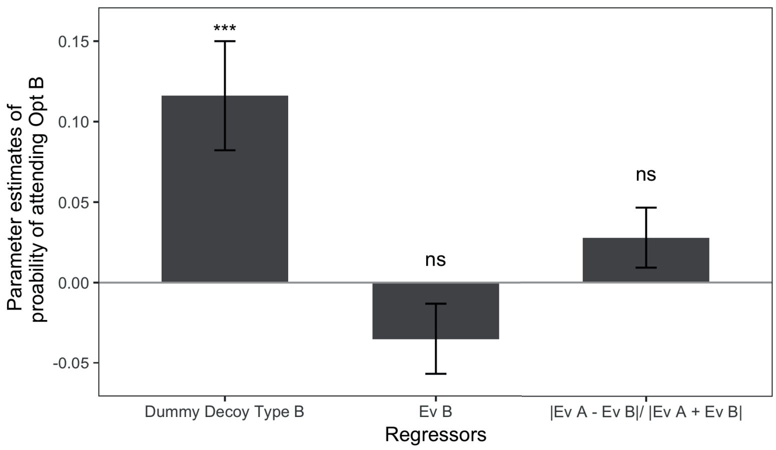


(b)

Figure B.11: a) Mixed-effect logistic regression model with random effects for subject-specific and trial-specific constants and slopes of the probability of attending the money attribute of option A, B and D, as a function of the distances (see Model summary for the definition of distance and proximity) from others (A,B,D) in the money attribute. b) Mixed-effect logistic regression model with random effects for subject-specific and trial-specific constants and slopes of the probability of attending the probability attribute of option A, B and D, as a function of the distances (see Model summary in the supplementary material for the definition of distance and proximity) from others (A,B,D) in the money attribute.



(a)



(b)

Figure B.12: a) Mixed-effect logistic regression model with random effects for subject-specific and trial-specific constants and slopes of the probability of attending option B as a function of the expected value of B and the normalized distance from A. b) Mixed-effect logistic regression model with random effects for subject-specific and trial-specific constants and slopes of the probability of attending option B as a function of the expected value of B and the normalized distance from A. Note, similar results hold when analysing the probability of attending option A or the probability of attending the decoy. Also note, the interactions are not displayed in the plots but have been included in the regressions.

B.7 Supplementary Material

B.7.1 Models summary

Multialternative decision theory (MDFT) and multiattribute leaky competing accumulators (MLCA)

The MDFT (Hotaling et al., 2010; Roe et al., 2001) and the MLCA models are both sequential sampling type of models, i.e., they assume that decisions are made by accumulating evaluations (or evidence) until a threshold is reached, at which point the choice is made. Both models do not assume a prominent role of attention in the attraction effect which is instead explained by a distance-dependent lateral inhibition in the MDFT and by an asymmetric value function that accounts for loss aversion in the MLCA. However, other decoy effects like the similarity effect are explained by attention mechanisms in the models. Thus, we review the attentional aspect of the models and investigate possible testable predictions about attention which are independent of the decoy effect under consideration, so that we can test them in our data. Both models can be formalized as follows.

$$A_i(t+1) = \lambda A_i(t) + (1-\lambda)\left[I_i(t) - \beta \sum_{j \neq i} A_j(t) + E(t)\right], \tag{SB1}$$

where $A_i(t)$ is evidence accumulated at time t for option i in the activated attribute, λ is the neural decay constant, β is the global inhibition parameter, $E(t)$ corresponds to a normally distributed noise term, and $I_i(t)$ is the input value which depends on the distance between options in the activated attributes. The input value function can be written as

$$I_i(t) = \sum_{j \neq i} V(d_{ij}), \tag{SB2}$$

where V is the value function and d_{ij} is the distance between option i and j in the activated attribute. The evidence is accumulated in one specific attribute, i.e., one attribute is activated, whenever that attribute is attended, then the input function I_i for each option i is calculated as a function of the distance between options in that specific attribute. Thus, the models assume that a comparison process between options in each attribute is performed and each attribute has some probability of being selected at each time step.

Associations and accumulation of preferences model (AAM)

The associations and accumulation model (AAM) uses a connections network model of the decision process, which assumes an association between choice task and attribute accessibility within a stochastic sequential-sampling accumulation framework. This model assumes that the *associative* connection between an option and an attribute is proportional to the presence of the attribute in the option. In particular, an option is strongly associated with an attribute if the option has a large amount of the attribute. Namely, in this model it is assumed that the associative connection between an alternative and an attribute is equal to the amount of the attribute in the alternative.

In mathematical terms, the accessibility of an attribute i, at any time period depends on the value of the attribute of all the options in the choice set.

$$a_i = a_0 + \sum_X s_x V X_i, \tag{SB3}$$

where s_x is the positive activation given to available and salient options, VX_i is the value (or association value) of option X in attribute i and a_0 is the constant activation identical to all attributes that serves to moderate the strength of the proposed associative biases.

The probability of the attribute being attended to is determined by the attribute's accessibility. This can be written as

$$P(\text{Attending attribute} i) = \frac{a_i}{\sum_j a_j}. \tag{SB4}$$

If an attribute i is sampled at any time t, the preference state for option X at time t can be written

$$P_X(t+1) = dP_X(t) + U_i(X) + \varepsilon(t), \tag{SB5}$$

where $U_i(X)$ is the value function that takes as an input VX_i – i.e., the value of the attribute i of option X. A decision is made when the preference crosses a threshold.

Multiaternative Decision by Sampling (MDbS) model

In the MDbS model, evidence is accumulated from a series of ordinal comparisons of pairs of attribute values. For example, in evaluating the price of Car A, no matter the source of the comparison attribute, if the price of Car A is preferable in the pairwise comparison, one unit of evidence is accumulated toward deciding on Car A. This pairwise comparison is considered ordinal, in the sense that evidence is increased one single unit amount regardless of how large the difference is.

The MDbS model is guided by three main constraints. First, the existing literature shows that, in multialternative decision, people's attention fluctuates between pairs of alternatives on single attributes at one time. The second constraint is that more similar alternatives receive more attention. Thus, in the MDbS model, more similar alternatives are more likely to be selected for comparison. Third, the distribution of time taken to make a decision (response time) is generally positively skewed and, toward the end of a decision, the decision makers attend more to the alternative which they are going to choose (the gaze cascade effect).

In MDbS, the probability of evaluating the value of option A on attribute i is proportional to the similarity to the other attribute values in working memory:

$$P(\text{Evaluate}A_i) \propto \sum_{X_i \neq A_i} \exp(-\alpha D_{(A_i,X_i)}), \tag{SB6}$$

where A_i is the attribute value for alternative A on dimension i, $D_{(A_i,X_i)} = \frac{|A_i - X_i|}{|X_i|}$.

In the MDbS model, the rate at which the evidence in favor of alternative A is accumulated depends on the probability of option A is evaluated in attribute i and the probability that A_i wins the comparison with other alternatives in attribute i. Thus,

$$
\begin{aligned}
&P(\text{Evidence accumulated towards} A) = && \text{(SB7)}\\
&\sum_{i \in D} P(\text{Evaluate } A_i) P(A_i \text{ wins a comparison}) = \\
&\sum_{i \in D} P(\text{Evaluate } A_i) \sum_{X_i \neq A_i} P(A_i \text{ is compared against } X_i) P(A_i \text{ is favored against } X_i).
\end{aligned}
$$

Hence, the MDbS predicts that the probability of attending one attribute of one option inversely depends on the distances from other options in the same attribute. Also, the model predicts that more similar options are attended more often, meaning more transitions between similar options than non-similar options. Concerning the attraction effect, this should translate in more transitions Target-Decoy than Target-Competitor.

Mutual Inhibition value-based attention capture (MIVAC) model

In re-analyising data from Chau et al. 2014, they found more evidence for the hypothesis that options seem to capture attention proportional to their value. The MIVAC is an another example of sequential sampling model with leakage and inhibition terms.

Contrary to the MDFT or the LCA, the input function does not depend on the distance between option in one specific attribute, but it is dependent on the integrated values of the options. The extra feature of this model is the probability of attending an option which gives a constant boost to the accumulator of the attended option. Formally,

$$\mathbf{A}_{t+1} = \mathbf{S}\mathbf{A}_t + \mathbf{I}_t + \mathbf{E}_t, \tag{SB8}$$

Where $\mathbf{S}$ represents an n by n leakage and inhibition matrix, $\mathbf{I}_t$ represents an n by t input matrix and $\mathbf{E}_t$ n by t noise matrix. The probability of attending an option is given by

$$F_i = \frac{e^{\alpha \times nEV(i)}}{\sum_j e^{\alpha \times nEV(j)}}, \tag{SB9}$$

where α is a free parameter and $nEv(i)$ is the normalized expected value of option i - normalization was used to avoid that value-based attentional capture effects depend too much on the sum of all values. Every time step an option (A for example) is attended the input to the accumulator for that option is

$$I_{A,t} = EV(A) + \beta, \tag{SB10}$$

and when A is not attended the input reads

$$I_{A,t} = EV(A), \tag{SB11}$$

where β is a free parameter representing the attention-based enhancement of accumulation. However, this model cannot account for the attraction effect in this setting. To explain the attraction you need to make the input to the accumulator as a function of comparisons of one option with all the other options.

The multialternative attentional drift diffusion model (MaDDM)

The MaDDM is another example of sequential sampling model without leaky integration or inhibition term. The key feature of this model is that the evidence accumulated depends on each time step on the attended option and the subjective value of the unattended options are discounted by a factor which underweight their value. The model can be formalized as follows.

$$A_i(t+1) = A_i(t) + dI_i(t) + E(t), \tag{SB12}$$

where $A_i(t)$ is evidence accumulated at time t for option i, d is the constant speed of accumulation and E_t corresponds to a normally distributed noise term. The input function for this model can be written as

$$I_i(t) = \theta_i V_i, \tag{SB13}$$

where V_i is the subjective value of option i and θ_i is equal 1 when option i is attended and equal to a constant between 0 and 1 when option i is not attended.

The model does not make any predictions in term of probability of attending one or the other option, but the eye-gaze data must be given exogenously to the model.

Pairwise Attribute Normalization

Pairwise Attribute Normalization (Landry and Webb, 2017) explains context-dependent effects with a model that makes two assumptions. First, information is normalized within each attribute dimension, and, second, normalization is pairwise in the sense that each pair of alternatives is compared separately within each attribute dimension. The model is the following.

"Our model considers a consumer who faces a finite choice set, X. Each alternative $x \in X$ is defined with respect to n > 0 attributes, where $x_i \geq 0$ is the unnormalized input value for alternative x on attribute $i \geq n$. Letting $C_x \in X/\{x\}$ denote the set of all alternatives in X besides x, the consumer's overall valuation of $x \in X$ is then given by the pairwise attribute normalization value function as follows:

$$v(x;C_x) = \sum_{y \in C_x} \sum_{i=1}^{n} \frac{x_i}{x_i + y_i}. \tag{SB14}$$

We use (SB14) as the basis of our descriptive choice model in that the consumer is presumed *to choose $x \in X$ if $v(x;C_x) > v(y;C_y)$ for all $y \in C_x$. More generally, we will say that x is "preferred" to y given X if $v(x;C_x) > v(y;C_y)$"*

Thus, this model makes one simple testable prediction. Fixing the normalized value difference, increasing or decreasing the absolute value of the options' attributes should not have an effect on choices. We check this prediction in our data. We first calculate the normalized value difference $v(x;C_x) - v(y;C_y)$ for both attributes as follows.

$$v_m(A;C_x) = v_m(A;B) + v_m(A;D) = \frac{VA_m}{VA_m + VB_m} + \frac{VA_m}{VA_m + VD_m} \tag{SB15}$$

$$v_p(A;C_x) = v_p(A;B) + v_p(A;D) = \frac{VA_p}{VA_p + VB_p} + \frac{VA_p}{VA_p + VD_p}. \tag{SB16}$$

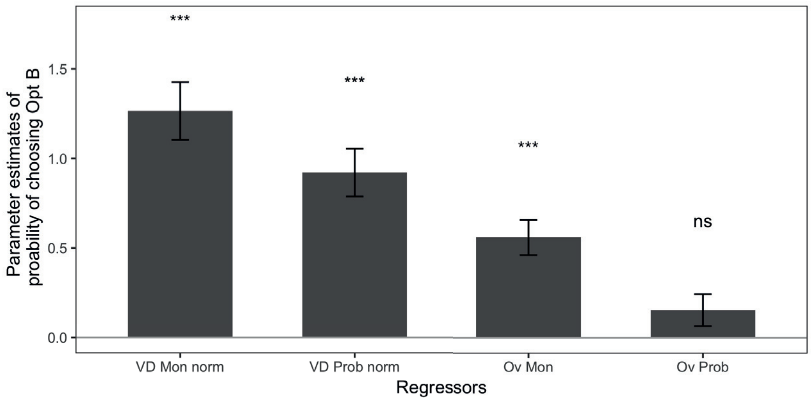
Finally, we define the overall value in the money attribute and the overall value in the probability attribute as follows

$$OV_m = VA_m + VB_m + VD_m$$

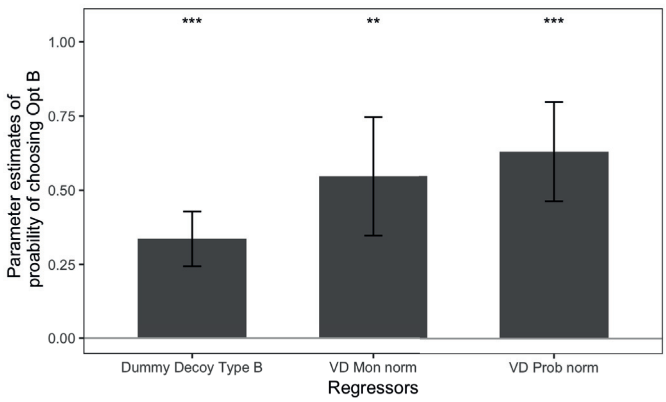
$$OV_p = VA_p + VB_p + VD_p$$

Figure SB1 shows the logistic regression model of the probability of choosing option B as a function of the normalized values calculated as above. We can notice that the overall value, i.e., the sum of all options' money, has an impact on choice frequency which goes against pairwise normalization theory. Moreover, if the pairwise Attribute Normalization would be able to explain the decoy effect, then the normalized value differences of the two attributes should be able to explain all the variability in the decoy effect. Figure SB1b shows an effect of the decoy type dummy on the probability of choosing option B.

B.7.2 Figures and Tables

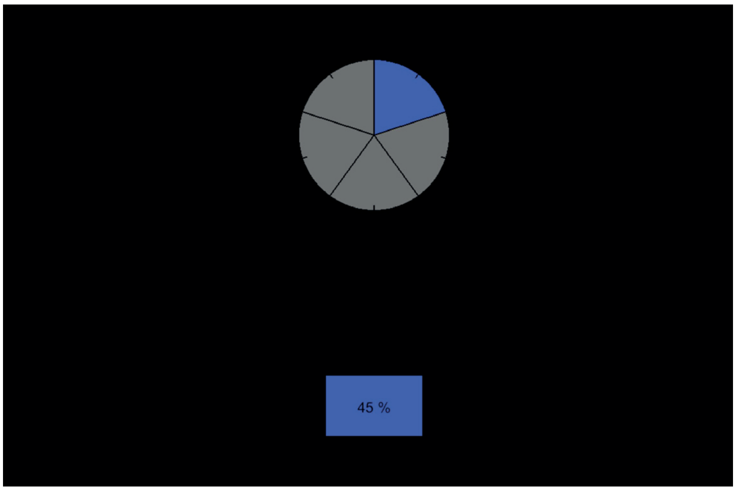


(a)

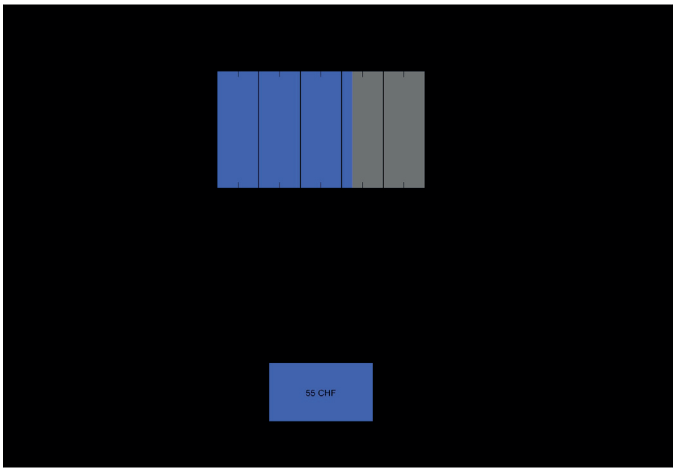


(b)

Figure SB1: a) Mixed-effect logistic regression model with random effects for subject-specific and trial-specific constants and slopes of the probability of choosing option B as a function of the normalized value difference in money (VD mon norm), the normalized value difference in probabilities (VD prob norm), the overall value in money and probability (OV mon and OV prob, respectively). b) Mixed-effect logistic regression model with random effects for subject-specific and trial-specific constants and slopes of the probability of choosing option B as a function of the normalized value difference in money (VD mon norm), the normalized value difference in probabilities (VD prob norm), and a dummy variable of decoy type B (baseline = decoy type A). Note, the interactions are not displayed in the plots but have been included in the regressions.



(a)



(b)

Figure SB2: Training phase. (a) Probability training phase. Subjects had to specify the probability represented by the colour-filled area of the pie-chart. (b) Money training phase. Subjects had to specify the amount of money represented by the colour-filled area of the rectangle.

	Dependent variable:		
	Option B chosen		
	(1)	(2)	(3)
Constant	0.489	0.738^{*}	0.641 .
	(0.342)	(0.329)	(0.378)
Decoy Type B	0.491^{***}		0.292 .
	(0.104)		(0.150)
EvB-EvA	0.454^{***}	0.457^{***}	0.534^{***}
	(0.086)	(0.074)	(0.105)
OV	0.154	0.117	0.295 .
	(0.126)	(0.118)	(0.154)
% Fix Opt B		1.513^{***}	1.606^{***}
		(0.132)	(0.155)
% Fix Decoy		0.579^{***}	0.393^{***}
		(0.081)	(0.103)
Decoy Type B × EvB-EvA	0.031		−0.090
	(0.101)		(0.130)
Decoy Type B × OV	-0.267^{**}		-0.340^{*}
	(0.096)		(0.133)
Decoy Type B × % Fix Opt B			−0.022
			(0.177)
Decoy Type B × % Fix Decoy			0.345^{*}
			(0.156)
% Fix Opt B × EvB-EvA		0.230^{*}	0.330^{**}
		(0.090)	(0.126)
% Fix Opt B × OV		-0.212^{*}	−0.092
		(0.091)	(0.136)
% Fix Decoy × EvB-EvA		−0.044	0.023
		(0.075)	(0.114)

	(1)	(2)	(3)
% Fix Decoy × OV		−0.061	−0.016
		(0.073)	(0.110)
Decoy Type B × % Fix Opt B × EvB-EvA			−0.162
			(0.185)
Decoy Type B × % Fix Opt B × OV			−0.221
			(0.196)
Decoy Type B × % Fix Opt Decoy × EvB-EvA			−0.088
			(0.145)
Decoy Type B × % Fix Opt Decoy × OV			−0.025
			(0.143)
Log Likelihood	−2,570	−2,136	−2,070
Akaike Inf. Crit.	5,194	4,379	4,518
Bayesian Inf. Crit.	5,373	4,737	5,770
Pseudo-R^2	0.32	0.44	0.46
Note:	. $p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$		

Table SB1: Three logistic mixed-effects regressions with random effects for subject-specific constants and slopes of the probability of choosing option B. Dependent variable equals 1 if option B is chosen and 0 otherwise. (1) The regressors for model 1 are the following: a dummy variable for decoy type B (baseline decoy type A), expected value difference Ev B – Ev A, overall value (OV = Ev B + Ev A), and all possible interactions. (2) The regressors for model 2 are the following: expected value difference Ev B – Ev A, overall value (OV = Ev B + Ev A), proportion of fixation time to option B (% Fix Opt B), proportion of fixation to the decoy (% Fix Decoy) and all possible interactions. (3) Model 3 combines all the regressors of model 1 and 2 plus all possible interactions.

Note, all independent variables are standardize, thus it is possible to compare the beta magnitude.

	Dependent variable:			
	Proportion of Fix Time option A		Proportion of Fix Time option B	
	(1)	(2)	(3)	(4)
Constant	0.395***	0.404***	0.382***	0.382***
	(0.008)	(0.008)	(0.009)	(0.009)
Decoy Type B	−0.028***	−0.034***	0.015*	0.017*
	(0.007)	(0.007)	(0.007)	(0.007)
EvB-EvA	−0.004		0.004	
	(0.003)		(0.003)	
Decoy Type B × EvB-EvA	−0.005		0.006	
	(0.004)		(0.004)	
EvA		0.007		
		(0.003)		
Decoy Type B × EvA		0.007		
		(0.004)		
EvB				−0.010
				(0.003)
Decoy Type B × EvB				0.002
				(0.004)
Observations	5,407	5,407	5,423	5,423
Log Likelihood	3,282.720	3,291.892	3,306.167	3,311.372
Akaike Inf. Crit.	−6,535.439	−6,553.785	−6,582.335	−6,592.744
Bayesian Inf. Crit.	−6,436.508	−6,454.853	−6,483.359	−6,493.768
Note:				. p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table SB2: Three linear mixed-effects regressions with random effects for subject-specific constants and slopes of the proportion of fixation time to option A - models (1) and (2) - and of the proportion of fixation time to option B - models (3) and (4). The regressors of models (1) and (3) are the following: a dummy variable for decoy type B (baseline decoy type A), the expected value difference Ev B – Ev A, and their interactions. The regressors of models (2) and (4) are the following: a dummy variable for decoy type B (baseline decoy type A), the expected value of option A (EvA) or the expected value of option B in model (1) or (3) respectively, and their interactions.

	Dependent variable:
	Count of Number of transitions
Constant in Baseline i.e., Type Transition: TARGET - DECOY	0.983^{***} (0.051)
Type Transition: COMPETITOR - DECOY	-0.524^{***} (0.040)
Type Transition: TARGET - COMPETITOR	0.169^{***} (0.043)
OV baseline	-0.074^{**} (0.026)
| EV target - EV competitor | baseline	0.053 (0.051)
Type Transition: COMPETITOR - DECOY $\times$ OV	-0.025 (0.026)
Type Transition: TARGET - COMPETITOR $\times$ OV	0.033 (0.026)
Type Transition: COMPETITOR - DECOY $\times$ | EV target - EV competitor |	-0.038 (0.045)
Type Transition: TARGET - COMPETITOR $\times$ | EV target - EV competitor |	-0.091 (0.061)
Note:	. $p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table SB3: A poisson mixed-effects regression with random effects for subject-specific constants and slopes of the count of number of transitions' gazes from one option to the other per trials. The regressors of the model are the following: a categorical variable for the type of transition (the baseline is transition type TARGET-DECOY) which is 0 if the count is for transitions from target to decoy or vice-versa, it is 1 for the count of transitions between competitor and decoy (transition type COMPETITOR-DECOY) and it is 2 for the count of transitions between target and competitor (transition type TARGET-COMPETITOR); the overall value (OV = EV target + EV competitor), the distance in expected value between target and competitor (absolute value of EV target – EV competitor), and all possible interactions.

	Dependent variable:
	Decoy Effect
Constant	0.089^{***}
	(0.012)
% Fix Target - % Fix Competitor	0.073^{***}
	(0.018)
% Fix Decoy	-0.020
	(0.014)
(% Fix Target - % Fix Competitor) × % Fix Decoy	0.037^{***}
	(0.008)
Note:	. p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table SB4: Linear regression model on the mean decoy effect, i.e., difference in probabilities between choosing the target and choosing the competitor, as a function of the mean proportion of fixation time difference between target and competitor (% Fix Target - % Fix Competitor), the mean proportion of fixation time to the decoy (% Fix Decoy) and their interactions.

Appendix C: Study 3

Simple method to fit hierarchical Bayesian piecewise time-constant drift diffusion models.

Gaia Lombardi and Todd Hare

University of Zürich

Department of Economics

Zürich Center for Neuroeconomics (ZNE)

C.1 Abstract

Noisy accumulation of evidence to a decision bound has been widely used in psychology and neuroscience to model decision-making process for value-based and perceptual decisions - e.g. drift diffusion model (DDM), race models. For mainly simplicity reasons in computational setups, most of the models in literature have assumed a time-constant drift rate (i.e. constant speed of the accumulation of the evidence in time) as the standard DDM or the leaky competing accumulator model (LCA). Additionally, the rise of hierarchical Bayesian estimation methodologies have further penalised the use of time-varying drift rate models as they require an analytical solution for the model which is not always possible to mathematically derive. Even though there are attempts to approximate the analytical solution for some type of time-varying drift rate DDMs, it is theoretically and computationally challenging to estimate and implement such formulations.

Here, we developed a theoretically and practically simple method to estimate drift diffusion models for which the drift rate is piecewise time-constant. Not only our method is very practical to implement and estimate, further it allows the use of pre-existing hierarchical Bayesian methodologies already developed for the standard DDM. To demonstrate the efficacy and the benefit of this method, we apply it to two practical examples, the attentional drift diffusion model (aDDM) and the time-varying sequential sampling (tSSM). We derive

the mathematical formulations of the method for these two examples, and we apply them to simulated and real data sets. Our results show that the method allows us to quickly and accurately recover parameters from simulations and fit real data set in a hierarchical Bayesian fashion.

C.2 Introduction

Noisy accumulation of evidence to a decision bound has proven to be an important and highly influential concept within models of value-based and perceptual decisions in the social and biological sciences. This process is at the core of widely used sequential sampling models such as the drift diffusion model (DDM; Ratcliff, 1978; Smith and Ratcliff 2004; Ratcliff et al. 2001; White et al. 2010), decision field theory (MDFT; Hotaling et al., 2010; Roe et al., 2001), the leaky competing accumulators model (LCA; Usher and McClelland, 2004), the association and accumulation model (AAM; Bathia 2013), the linear ballistic accumulator (LBA; Brown and Heathcote, 2008), and several. To date, most examples of this class of model that have been fit to empirical data have assumed a constant evidence accumulation or drift rate throughout the choice process (e.g. the standard DDM or the LCA). One important reason for maintaining this assumption has been the practical considerations relating to the computational complexity and time required to fit models that relax this assumption. The rise of hierarchical Bayesian estimation methodologies (Vandekerckhove, Tuerlinckx, and Lee, 2011; Wiecki, Sofer, and Frank 2013; Wabersich and Vandekerckhove 2014) may have further dissuaded researchers from using time-varying drift rate models because Bayesian estimation using popular Markov-chain Monte Carlo (MCMC) methods generally require an analytical solution for equations within the model in order to generate the posterior distribution. Bayesian data analytic methods have at least two benefits for the decision sciences. First, Bayesian methods allow inference on the entire distribution of the parameters rather than just the most likely estimate as in methodologies based on maximum likelihood (MLE). These full distributions of parameters are useful because they provide a measure of the uncertainty in the parameters' values. Second, hierarchical structures are efficiently embedded in the Bayesian frameworks. Standard modelling methodologies (Myung 2003; Turner and Sederberg 2012; Turner and Sederberg 2014) have either to assume complete independence in the data structure (i.e., as in single-subject MLE) or complete pooling (i.e., fitting averaged data and implicitly assuming that all participants are identical), whereas hierarchical Bayesian modelling implements a compromise between the two strategies allowing group and subject parameters to be estimated simultaneously at different hierarchical levels, and thus also providing robust estimates of a model's free parameters without ignoring or over-weighting individual differences. Namely, they can allow for the more complete theoretical explanation of data from a single task, letting different people use different cognitive processes, or letting the same people use different processes at different times. However, as mentioned above, efficiently generating the posterior distribution of the parameters in hierarchical Bayesian methods requires an analytical solution for equations within the model framework. Analytical solutions are not always possible to mathematically derive for models with time-varying evidence accumulation (i.e. drift) rates. Although there are attempts to approximate the analytical solution for some types of time-varying drift rate DDMs (Srivastavaa et al. 2015), it often remains theoretically and computationally challenging to estimate and implement such formulations. Despite the added complexity of their estimation, time-varying drift rate models are important because there are many types of decision strategies and mechanisms that cannot be modeled with a fixed drift rate. In fact, there are several notable examples in literature in which sequential sampling models have a time-varying drift rate, such as the attentional drift diffusion model (aDDM; Krajbich and Rangel, 2011; Krajbich et al., 2010), the relative-starting-time drift diffusion model (stDDM, in preparation), the dual-stage two-phase (DSTP) model (Hubneret al. 2010), the piecewise linear ballistic accumulator (pLBA; Trueblood and Holmes 2016; Holmes and Trueblood 2018) and others. Critically, all of the models listed above share the fact that they do not have completely time-dependent drift rates in the sense of a linear or quadratic relation to time. Instead the drift rate is time-constant in fixed or alternating intervals, i.e. it is piecewise time-constant, in all of these models. Diffusion models with piecewise time-constant drift rates (pcDDM) can be approximated with a standard drift diffusion model (DDM) for which the drift rate is constant in time. Here, we show how a simple, yet powerful, piecewise constant approximation method allows for the use of existing hierarchical Bayesian estimation tools that were developed for the standard DDM to approximate the time-varying drift rate of pcDDMs. The advantage of using well-establish standard DDM methodologies can be found in the faster speed of estimation and in the simplicity of implementation when comparing to other methods such as the martingales method (Srivastavaa et al. 2015) or the probability density approximation (PDA) method (Turner and Sederberg 2014; Holmes 2016; Holmes and Trueblood 2018).

In this paper, we illustrate how to transform a pcDDM into a standard DDM with a fully time-constant drift rate. We will show that a piecewise time-constant drift rate can be expressed as a constant drift that is a function of the first passage time - i.e reaction time - and the times for each constant interval. We explain in detail how to derive the drift rate function from a discrete version of the pcDDM - i.e. bounded accumulation series model or Euler method - that is a numerical approximation of the continuous version - i.e. stochastic differential equation (SDE) process. Then, we demonstrate how to substitute the constant drift rate of the continuous DDM with this derived drift function in order to obtain a pcDDM in a continuous formulation that can be solved as a standard DDM. We also walk through two concrete examples of how to apply this methodology. We derive the piecewise time-constant drift for the aDDM and for the stDDM, and implement these drifts in a hierarchical Bayesian DDM implemented in R and Jags. We then show the recovery fitting analysis of the two models to demonstrate that 1) we correctly recover the time-varying parameters through this reparameterization of the standard DDM, and 2) that the other DDM parameters are not affected by the inclusion of these extra parameters for the drift rate. Lastly, we will fit pcDDM models using the methods we've outlined to two experimental data sets, and show goodness of fit measures for the experimental reaction time and choice distributions.

C.3 Results

We first introduce two different formulations of the standard DDM, the continuous version and the discrete approximation of the continuous version - i.e. Euler method. We show how it is possible to derive a closed-form solution for the Euler method in the DDM and apply the same procedure to the pcDDM. From this closed-form solution is possible to derive the piece-wise time-constant drift rate of the pcDDM and express it as a function of the reaction time and the interval's times, to obtain a constant drift rate similar to the one of the DDM. We show this procedure in general terms and then we apply it to two example models, the aDDM and stDDM.

Standard drift diffusion model (DDM)

A standard drift diffusion model is commonly expressed in two different forms in the literature: 1) as a stochastic differential equation (SDE), or 2) as a bounded accumulation series model. The first form of the model makes use of stochastic accumulation in a continuous time framework - i.e. SDE, whereas the latter formalises the accumulation process as a discrete bounded accumulation series which is a numerical approximation - i.e. Euler method - of the SDE form of the model. The two formulations can be expressed as following.

$$\begin{aligned} &\mathrm{d}x(t) = \mu \ \mathrm{d}t + \sigma \ \mathrm{d}W(t) \qquad x(0) = x_0, \\ &\tau = \inf\{t > 0 | x(t) \notin (-B, B)\} \end{aligned} \tag{C.1}$$

$$\begin{aligned} &x_{t_i} = x_{t_{i-1}} + \mu(t_i - t_{i-1}) + \varepsilon_{t_i} \qquad x_{t_0} = x_0, \\ &\tau = t_N = \min\{t_i, \forall i \in \mathbb{N} | \ |x_{t_i}| \geq B\} \end{aligned} \tag{C.2}$$

where $x(t)$ (x_{t_i}) is the evidence accumulated at time t (t_i), τ the first passage time (or reaction time) when the evidence crosses a certain threshold B, μ is the drift rate or speed of the accumulation process, σ is the diffusion rate or standard deviation of the brownian motion, x_0 is the initial bias and $\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$.

In equation C.1, the evidence $x(t)$ evolves in time according to a biased random walk with Gaussian increments, i.e. $\mathrm{d}x(t) \sim \mathrm{N}(\mu\ \mathrm{d}t, \sigma^2\ \mathrm{d}t)$, and the analytical solution for the distribution of the reaction time τ can be easily derived when drift rate μ and diffusion rate σ are constant in time (see Navarro and Fuss, 2009, Fast and accurate calculations for first-passage times in Wiener diffusion models).

Concerning the second formulation of the DDM, equation C.2, the decision variable is modelled as a discrete stochastic evolution towards one or the other decision threshold -B or B. Although this formulation of the model is mostly used to generate simulations for the evidence accumulation process and response times, equation C.2 has a closed-form solution that can be derived by simply solving the series recursively until the first passage time τ. We will derive the closed-form solution in details for equation C.2 in the next paragraph.

Closed-form solution of the bounded accumulation series

Let us assume for simplicity that the time step is equal to 1 - i.e. $t_i - t_{i-1} = 1$. If the first passage time step N is known, then the bounded accumulation series can be solved recursively for each time step from N to 0, and a closed solution for C.2 can be derived as follows.

$$
\begin{aligned}
x_{t_N} &= x_{t_{N-1}} + \mu + \varepsilon_{t_N} \\
x_{t_{N-1}} &= x_{t_{N-2}} + \mu + \varepsilon_{t_{N-1}} \\
x_{t_{N-2}} &= x_{t_{N-3}} + \mu + \varepsilon_{t_{N-2}} \\
&\dots \\
x_{t_N} &= x_{t_{N-2}} + \mu + \varepsilon_{t_{N-1}} + \mu + \varepsilon_{t_N} \\
x_{t_N} &= x_{t_{N-3}} + \mu + \varepsilon_{t_{N-2}} + \mu + \varepsilon_{t_{N-1}} + \mu + \varepsilon_{t_N} \\
&\dots \\
x_{t_N} &= x_{t_0} + \sum_{i=1}^{N} \mu + \sum_{i=1}^{N} \varepsilon_{t_i}
\end{aligned}
\tag{C.3}
$$

Further, we can resolve the first sum and divide both sides of the equation by a positive constant and without changing the result of the accumulation process and the first passage time.

$$
\begin{aligned}
\frac{1}{N} x_{t_N} &= \frac{1}{N} x_{t_0} + \frac{1}{N} N\mu + \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{t_i} \\
\widetilde{x}_{t_N} &= \widetilde{x}_{t_0} + \mu + \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{t_i}
\end{aligned}
\tag{C.4}
$$

where $\widetilde{x}_{t_i} = \frac{1}{N} x_{t_i}$, $\forall i \in \mathbb{N}$ and $\tau = t_N = \min\{t_i, \forall i \in \mathbb{N} |\ |x_{t_i}| \geq B\} = \min\left\{t_i, \forall i \in \mathbb{N} |\ |\widetilde{x}_{t_i}| \geq \frac{B}{N}\right\}$.

We showed that knowing the first passage time, N, we can derive the closed-form solution for the discrete series of the DDM. We will show in the next section how this solution changes when the drift rate μ is not constant in time, but is instead piece-wise constant.

General piece-wise time-constant DDM (pcDDM)

As noted in the introduction section of the paper, some of the most prominent time-varying DDMs found in literature, e.g. the aDDM, do not have fully time-dependent drift rates, but rather have drift rates that are constant over discrete intervals. The continuous SDE of such a model (pcDDM) can be written as follows.

$$\begin{aligned}
&\mathrm{d}x(t) = \mu(t)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t) \qquad x(0) = x_0, \\
&\mu(t) = \mu_i, \quad \text{for } t_i \leq t < t_{i+1} \\
&\tau = \inf\{t > 0 | x(t) \notin (-B, B)\}
\end{aligned} \tag{C.5}$$

and the corresponding numerical approximation - i.e bounded accumulation series formulation, is

$$\begin{aligned}
&x_{t_i} = x_{t_{i-1}} + \mu_{t_i}(t_i - t_{i-1}) + \varepsilon_{t_i} \qquad x_{t_0} = x_0, \\
&\mu_{t_i} = \mu_j, \quad \text{for } s_j \leq t_i < s_{j+1} \\
&\tau = t_N = \min\{t_i, \forall i \in \mathbb{N}|\ |x_{t_i}| \geq B\}
\end{aligned} \tag{C.6}$$

where s_j is jth time step for which $\mu_{s_j} \neq \mu_{s_{j+1}}$, and $\mu_{t_i} = \mu_{t_{i+k}} = \mu_j$ for $t_i,\ t_{i+k} \in [s_j, s_{j+1})$, $\forall j \in \mathbb{N}$, $\forall k \in \mathbb{N}$.

The intuition behind our method for fitting such pcDDMs is to combine the piece-wise constant drift rates, $\mu_{s_1} : \mu_{s_n}$, into a single constant drift rate, μ_{ave}, that is equal to the weighted average of $\mu_{s_1} : \mu_{s_n}$ as in equation C.4. We can then use this μ_{ave} for estimating the analytical solution of the continuous pcDDM in equation C.5. Thus, we derive the closed-form solution for equation C.6 as we did in C.3 for the pcDDM, as follows.

$$
\begin{aligned}
x_{t_N} &= x_{t_{N-1}} + \mu_{t_N} + \varepsilon_{t_N} \\
&\quad x_{t_{N-1}} = x_{t_{N-2}} + \mu_{t_{N-1}} + \varepsilon_{t_{N-1}} \\
&\qquad x_{t_{N-2}} = x_{t_{N-3}} + \mu_{t_{N-2}} + \varepsilon_{t_{N-2}} \\
&\quad \dots \\
x_{t_N} &= x_{t_{N-2}} + \mu_{t_{N-1}} + \varepsilon_{t_{N-1}} + \mu_{t_N} + \varepsilon_{t_N} \\
&\quad x_{t_N} = x_{t_{N-3}} + \mu_{t_{N-2}} + \varepsilon_{t_{N-2}} + \mu_{t_{N-1}} + \varepsilon_{t_{N-1}} + \mu_{t_N} + \varepsilon_{t_N} \\
&\quad \dots \\
x_{t_N} &= x_{t_0} + \sum_{i=1}^{N} \mu_{t_i} + \sum_{i=1}^{N} \varepsilon_{t_i}
\end{aligned}
\tag{C.7}
$$

where we can write the sum of the drift rates as a function of the time intervals in which the drift rate is constant as follows.

$$
\sum_{i=1}^{N} \mu_{t_i} = \sum_{j=1}^{n-1} (s_{j+1} - s_j)\mu_j \tag{C.8}
$$

where $n-1$ is the number of times the drift rate changes, and it is constant (i.e. $\mu_{t_i} = \mu_j$) in $t_i \in [s_j, s_{j+1})$. Thus, we can divide both sides of the equation by the first passage time step N as we did in C.4 and derive the average constant drift rate for the pcDDM as follows.

$$
\begin{aligned}
\frac{1}{N} x_{t_N} &= \frac{1}{N} x_{t_0} + \frac{1}{N} \sum_{i=1}^{N} \mu_{t_i} + \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{t_i} \\
\widetilde{x}_{t_N} &= \widetilde{x}_{t_0} + \frac{1}{N} \sum_{j=1}^{n-1} (s_{j+1} - s_j)\mu_j + \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{t_i}
\end{aligned}
\tag{C.9}
$$

where $\frac{1}{N}\sum_{j=1}^{n-1}(s_{j+1} - s_j)\mu_j$ for $\mu_j \in \mathbb{R}$ is constant in $[s_j, s_{j+1})$. Then, if the first passage time is known, the continuous form of the pcDDM can be written as

$$\begin{aligned}
&\mathrm{d}x(t) = \mu \ \mathrm{d}t + \sigma \ \mathrm{d}W(t) \qquad x(0) = x_0, \\
&\mu = \frac{1}{\tau}\sum_{i=1}^{n-1}(t_{i+1} - t_i)\mu_i, \\
&\tau = \inf\{t > 0 | x(t) \notin (-B, B)\}
\end{aligned} \tag{C.10}$$

where μ is a function of first passage time τ, the number of intervals n and their duration $t_{i+1} - t_i$.

In the next section, we will show the drift rate formula μ for two specific examples of pcDDM, the aDDM and the stDDM.

aDDM

The aDDM model is an example of a modified version of a standard DDM for which the drift rate varies in time. The basic idea of the model is that the drift rate changes every time the decision maker visually fixates one option or another, assuming there are only two options in the choice set. The discrete bounded accumulation series of the aDDM writes

$$\begin{aligned}
&x_{t_i} = x_{t_{i-1}} + \mu_{t_i} + \varepsilon_{t_i} \\
&\mu_{t_i} = \begin{cases} \delta(V_A - \theta V_B) & \text{if option A is attended at time } t_i \\ \delta(\theta V_A - V_B) & \text{if option B is attended at time } t_i \end{cases} \\
&\tau = t_N = \min\{t_i, \forall i \in \mathbb{N} | \ |x_{t_i}| \geq B\},
\end{aligned} \tag{C.11}$$

where θ is the discount factor that penalises the value of the unattended option, δ is the drift constant parameter and V_A and V_B are the values of option A and B, respectively. The series can be solved as

$$
\begin{aligned}
x_{t_N} =& x_{t_0} + \sum_{i=1}^{N} \mu_{t_i} + \sum_{i=1}^{N} \varepsilon_{t_i} \\
\sum_{i=1}^{N} \mu_{t_i} =& \sum_{i=1}^{N_A} \delta(V_A - \theta V_B) + \sum_{i=1}^{N_B} \delta(\theta V_A - V_B) \\
\sum_{i=1}^{N} \mu_{t_i} =& \sum_{i=1}^{N_A} \delta V_A - \sum_{i=1}^{N_A} \delta\theta V_B + \sum_{i=1}^{N_B} \delta\theta V_A - \sum_{i=1}^{N_B} \delta V_B
\end{aligned}
\quad . \tag{C.12}
$$

As for the generalised pcDDM in C.9, we can resolve the first sum and divide both side of the equation by a positive constant without affecting the result of the accumulation process, as follows

$$
\begin{aligned}
\frac{1}{N} x_{t_N} =& \frac{1}{N} x_{t_0} + \frac{1}{N} N\mu + \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{t_i} \\
\widetilde{x}_{t_N} =& \widetilde{x}_{t_0} + \mu + \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{t_i} \\
\mu =& \frac{1}{N} \sum_{i=1}^{N} \mu_{t_i}
\end{aligned}
\tag{C.13}
$$

where $\widetilde{x}_{t_i} = \frac{1}{N} x_{t_i}$, $\forall i \in \mathbb{N}$ and $\tau = t_N = \min\{t_i, \forall i \in \mathbb{N} |\ |x_{t_i}| \geq B\} = \min\left\{t_i, \forall i \in \mathbb{N} |\ |\widetilde{x}_{t_i}| \geq \frac{B}{N}\right\}$. Thus, the drift rate for the aDDM results to be

$$
\begin{aligned}
\mu =& \frac{1}{N} \sum_{i=1}^{N_A} \delta V_A - \frac{1}{N} \sum_{i=1}^{N_A} \delta\theta V_B + \frac{1}{N} \sum_{i=1}^{N_B} \delta\theta V_A - \frac{1}{N} \sum_{i=1}^{N_B} dV_B \\
\mu =& \frac{N_A}{N} \delta(V_A - \theta V_B) + \frac{N_B}{N} \delta(\theta V_A - V_B)
\end{aligned}
\tag{C.14}
$$

where N_A is the number of time steps for which option A is attended and N_B is the number of time steps for which option B is attended. Finally, we can use the derived constant drift rate μ for the continuous formulation of the aDDM and rewrite the model as a standard DDM.

$$
\begin{aligned}
&\mathrm{d}x(t) = \mu \ \mathrm{d}t + \sigma \ \mathrm{d}W(t) \qquad x(0) = x_0, \\
&\mu = \frac{\tau_A}{\tau} \delta(V_A - \theta V_B) + \frac{\tau_B}{\tau} \delta(\theta V_A - V_B) \\
&\tau = \inf\{t > 0 | x(t) \notin (-B, B)\}
\end{aligned}
\tag{C.15}
$$

where τ_A and τ_B are the total fixation times to option A and option B, respectively.

stDDM

The relative-starting-time DDM (stDDM) is another example of a piece-wise constant DDM for multi-attribute choices. This formulation of the DDM includes a relative-starting-time parameter that allows for one attribute to begin being considered before another. Other than that additional parameter, the model is a basic drift diffusion model, in which the drift rate depends on the weighted value differences of two attributes. This model can be used for different types of options that have two or more attributes, for example lotteries with money and probabilities or food items with healthiness and tastiness attributes. The key idea of the model is that one of the two attributes might have some delay in entering the accumulation process relative to the other. For instance, a decision maker that has to choose between a salad and a chocolate cake might immediately think about the delicious taste of the cake compared to the salad, and only after some milliseconds start to consider the fact that the chocolate cake would be a far less healthy option.

The discrete bounded accumulation series of the stDDM can be formulated as follows.

$$\begin{aligned} x_{t_i} &= x_{t_{i-1}} + \mu_{t_i} + \varepsilon_{t_i} \\ \tau = t_N &= \min\{t_i, \forall i \in \mathbb{N} |\ |x_{t_i}| \geq B\} \\ \mu_{t_i} &= \begin{cases} w_T \mathrm{VD}_T & \text{if } s > 0 \ \wedge \ 0 < i < s \\ w_H \mathrm{VD}_H & \text{if } s < 0 \ \wedge \ 0 < i < |s| \\ w_H \mathrm{VD}_H + w_T \mathrm{VD}_T & \text{if } i > |s| \end{cases} \end{aligned} \tag{C.16}$$

where w_T is the weight given to the taste attribute, w_H the weight to the health attribute, VD_T and VD_H are the value differences in taste and in health respectively, s is the time step at which the health attribute comes into the accumulation process - $s > 0$ means taste is on from the beginning and health comes in at time t_s, $s < 0$ means that the health attribute is on the accumulation process from the beginning and taste comes in at time $t_{|s|}$.

Once again, we can solve the series recursively and obtain,

$$
\begin{aligned}
x_{t_N} &= x_{t_0} + \sum_{i=1}^{N} \mu_{t_i} + \sum_{i=1}^{N} \varepsilon_{t_i} \\
\sum_{i=1}^{N} \mu_{t_i} &= \begin{cases}
\sum_{i=1}^{s} w_T \mathrm{VD}_T + \sum_{i=s+1}^{N} \left(w_H \mathrm{VD}_H + w_T \mathrm{VD}_T\right) & \text{if } \; s > 0 \\
\sum_{i=1}^{|s|} w_H \mathrm{VD}_H + \sum_{i=|s|+1}^{N} \left(w_H \mathrm{VD}_H + w_T \mathrm{VD}_T\right) & \text{if } \; s < 0 \\
\sum_{i=1}^{N} w_T \mathrm{VD}_T & \text{if } \; s > 0 \;\wedge\; s > N \\
\sum_{i=1}^{N} w_H \mathrm{VD}_H & \text{if } \; s < 0 \;\wedge\; |s| > N
\end{cases}
\end{aligned}
\tag{C.17}
$$

As in the previous sections, we can divide both side of the equation by a positive constant without affecting the result of the accumulation process, as follows

$$
\begin{aligned}
\frac{1}{N} x_{t_N} &= \frac{1}{N} x_{t_0} + \frac{1}{N} N\mu + \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{t_i} \\
\widetilde{x}_{t_N} &= \widetilde{x}_{t_0} + \frac{1}{N} \sum_{i=1}^{N} \mu_{t_i} + \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{t_i}
\end{aligned}
\tag{C.18}
$$

where $\widetilde{x}_{t_i} = \frac{1}{N} x_{t_i}$, $\forall i \in \mathbb{N}$, and $\mu = \frac{1}{N} \sum_{i=1}^{N} \mu_{t_i}$. Then,

$$
\begin{aligned}
\mu &= \begin{cases}
\frac{1}{N} \sum_{i=1}^{s} w_T \mathrm{VD}_T + \frac{1}{N} \sum_{i=s+1}^{N} \left(w_H \mathrm{VD}_H + w_T \mathrm{VD}_T\right) & \text{if } \; s > 0 \;\wedge\; s < N \\
\frac{1}{N} \sum_{i=1}^{|s|} w_H \mathrm{VD}_H + \frac{1}{N} \sum_{i=|s|+1}^{N} \left(w_H \mathrm{VD}_H + w_T \mathrm{VD}_T\right) & \text{if } \; s < 0 \;\wedge\; |s| < N \\
\frac{1}{N} \sum_{i=1}^{N} w_T \mathrm{VD}_T & \text{if } \; s > 0 \;\wedge\; s > N \\
\frac{1}{N} \sum_{i=1}^{N} w_H \mathrm{VD}_H & \text{if } \; s < 0 \;\wedge\; |s| > N
\end{cases} \\
\mu &= \begin{cases}
\frac{s}{N} w_T \mathrm{VD}_T + \frac{N-s}{N} \left(w_H \mathrm{VD}_H + w_T \mathrm{VD}_T\right) & \text{if } \; s > 0 \;\wedge\; s < N \\
\frac{|s|}{N} w_H \mathrm{VD}_H + \frac{N-|s|}{N} \left(w_H \mathrm{VD}_H + w_T \mathrm{VD}_T\right) & \text{if } \; s < 0 \;\wedge\; |s| < N \\
w_T \mathrm{VD}_T & \text{if } \; s > 0 \;\wedge\; s > N \\
w_H \mathrm{VD}_H & \text{if } \; s < 0 \;\wedge\; |s| > N
\end{cases}
\end{aligned}
\tag{C.19}
$$

Finally, we can use the derived constant drift rate μ for the continuous formulation of the stDDM and rewrite the model as a standard DDM.

$$
\begin{aligned}
&\mathrm{d}x(t) = \mu \ \mathrm{d}t + \sigma \ \mathrm{d}W(t) \\
&\mu = \begin{cases} \frac{s}{\tau} w_T \mathrm{VD}_T + \frac{\tau - s}{\tau} (w_H \mathrm{VD}_H + w_T \mathrm{VD}_T) & \text{if } s > 0 \ \wedge \ s < \tau \\ \frac{|s|}{\tau} w_H \mathrm{VD}_H + \frac{\tau - |s|}{\tau} (w_H \mathrm{VD}_H + w_T \mathrm{VD}_T) & \text{if } s < 0 \ \wedge \ |s| < \tau \\ w_T \mathrm{VD}_T & \text{if } s > 0 \ \wedge \ s > \tau \\ w_H \mathrm{VD}_H & \text{if } s < 0 \ \wedge \ |s| > \tau \end{cases} \\
&\tau = \inf\{t > 0 | x(t) \notin (-B, B)\}
\end{aligned} \tag{C.20}
$$

In summary, we have shown that both the aDDM and the stDDM can be written as a standard DDM for which the drift rate is a function of the reaction time and the duration of the constant intervals (i.e. relative fixation time advantage or relative starting time advantage). The advantage of this formulation is that we have an analytical solution for the pcDDMs that coincides with the analytical solution of the DDM (Navarro and Fuss 2009), thus a hierarchical Bayesian approach can now be easily applied to this type of model. Here, we implement the piece-wise constant approximation method described above within a hierarchical Bayesian estimation framework giving us the hierarchical Bayesian attention drift diffusion model (HaDDM) and the hierarchical Bayesian relative-start-time DDM (HstDDM). We coded the HaDDM and the HstDDM in Jags using equations C.5 and C.20 above, and performed recovery fitting analysis in R (see the Method section for details and https://github.com/galombardi/ for the full code). Below we report and discuss tests of the parameter recovery accuracy for both models.

Recovery fitting analysis

To assess and demonstrate the validity of this method, we performed a parameter recovery analysis for the HaDDM and the HstDDM. We demonstrate that the piece-wise constant approximation method is able to correctly recover the parameters at the individual and at the group levels, and that the mean of the posterior distribution is informative of the real value of the generating parameter.

The aim of this recovery fitting analysis is twofold. Firstly, it is to show the ability of the method to recover different parameter combinations at the individual and the group level and, in particular, that the mean of the posterior distributions from the recover of each parameter is informative of the real generating values. Secondly, we want to show that this fitting method is able to recover different parameters for different individuals from the same population. Thus, we performed parameter recovery with two types of participant samples, A) samples of participants with homogeneous parameters (i.e. identical for all individuals) so that we could test several different combinations of parameters, and B) a participant sample with a heterogeneous distribution of parameters across individuals so that we could determine recovery accuracy at the individual level. In the homogeneous sample tests, for both the HaDDM and the HstDDM, the sets of parameters we tested varied around the best fitting parameter estimates from past fits to empirical data sets. For the heterogeneous samples, each of the parameters were randomly drawn from Gaussian distributions with a fixed mean and standard deviation. All simulations for both samples were performed with the bounded accumulation series version of the diffusion model for a fixed number of subjects, trials, and value differences that we took from existing empirical data sets (see Methods section for details). For all samples and versions of the diffusion model, we fitted a hierarchical Bayesian version of the model. As a brief reminder, an advantage of using hierarchical Bayesian estimation procedures is that we can directly estimate the posterior distributions of each parameter at the group and at the individual levels. Thus, hierarchical Bayesian fitting methods allow us to make inference about the population distributions of the parameters, while explicitly accounting for variability due to individual differences.

We will show in the next sections that the piece-wise constant approximation method was able to accurately estimate the parameters in recovery test for both the HaDDM and the HstDDM.

HaDDM

Fitting the HaDDM to groups of agents with homogeneous model parameters

We first tested parameter recovery accuracy for the HaDDM using a group of agents (N = 30 agents per group) that all made choices based on the same underlying model parameters. Specifically, we simulated the bounded accumulation series of the aDDM for several different groups while keeping the parameters combinations constant for all agents/subjects within a given group. The parameters' values were based on the range of parameters estimated for choices from real human subjects. The attention discount factor θ was drawn from the set $\{0.2, 0.4, 0.6, 0.8\}$, the drift constant δ from $\{6, 10, 14\}$ and the standard deviation of the noise σ from $\{0.3, 0.45, 0.6\}$, and then we simulated choices from samples having all the possible parameter combinations. Lastly, we fit the HaDDM using the piece-wise constant approximation method, and evaluated the recovered posterior distributions of the parameters for all simulated groups.

In figure C.1, we show the posterior distribution of the group-level mean of the discount parameter θ varying the drift scaling parameter δ and the standard deviation of the noise σ. Notably, the generating θ parameter was recovered well in all of the different groups. However, as expected, increasing the noise or decreasing the drift rate scaling parameter decreases the relative accuracy of the parameter recovery. We also confirmed that varying the attention discount factor θ does not influence the recovery accuracy of the other parameters. In Figures C.2 and C.3, we show the ability of the model to recover the generating δ and σ parameters respectively, varying the other two parameters. The piece-wise constant approximation method is able to accurately recover all the different parameters. However, it is worth noting that there is a systematic bias in the drift scaling parameter δ, which appears to be overestimated. A plausible explanation for this overestimation is the fact that the discrete bounded accumulation series is a numerical approximation of the continuous analytical solution of the model used to fit the parameters in the HaDDM. In other words, the bias we observe in the estimation may not be due to the piece-wise constant approximation method itself, but rather is the result of an approximation error between the discrete bounded accumulation series and the analytical solution of the DDM.

To test this hypothesis, we ran parameter recovery tests for the standard DDM using simulated choices and response times generated by either the bounded accumulation series in equation C.2 or from the analytical solution of equation C.1. We then estimated the best-fitting parameters for these data sets using the same hierarchical Bayesian method that we used for the standard DDM. As predicted, we find that, when choices are simulated with the bounded accumulated series, the drift rate scaling parameter in a standard DDM has the same systematic bias that we see for the aDDM (and in the stDDM reported in next section). Critically, this bias is absent when the choice response times are simulated with the analytical solution of the DDM as the generating model, appendix figure C.20. Thus, these results indicate that the bias stems from simulating the diffusion models with the bounded accumulation series, which is a numerical approximation of the continuous time model. If the evidence accumulation process in humans or other animals operates in continuous time, then there will be no bias. In any case, this bias is not problematic because it is a small, constant overestimation of the drift parameter. We will show in the next section that it does not significantly affect the goodness of fit of the response time distributions and it is even smaller at the individual than the group level (see for example figures C.18d and C.7).

Analyses of the differences between two sets of generating and recovered δ parameters showed that the piece-wise constant approximation method is able to fully distinguish between different generating δ parameters. Specifically, we calculated the posterior density of the difference between the drift scaling parameters, δ, from recovery fits for two different simulations, and compared it with the difference in the generating parameters. Figure C.4 shows that the recovered difference in δ estimates between simulations was highly accurate and unbiased. This is critical because in most cases researchers will be fitting piece-wise constant DDMs to compare across different individuals or experimental conditions.

Fitting the HaDDM to groups of agents with heterogeneous model parameters

It is also important to check whether our hierarchical Bayesian estimation method yields accurate estimates at the individual as well as the group level. When fitting choice data generated by humans or other animals, the groups or samples will not have homogeneous model parameters. Indeed, often it is the nature and magnitudes of differences between individuals in the sample that are the question of interest. Therefore, in a second set of parameter recovery analyses, we simulated data using different generating parameters for each agent/subject in the group. The generating parameters were randomly drawn from the gaussian distributions shown in Figure C.5a. Once again we were able to accurately recover the parameters, this time focusing on the individual level. Figure C.6, figure C.7 and figure C.8 show the recovery at the individual level for each subject for the θ, the δ and the σ parameters, respectively.

HstDDM

Fitting the HstDDM to groups of agents with homogeneous model parameters

We repeated the same procedures described for the aDDM to test parameter recovery from the stDDM. Thus, we simulated choice data from several combinations of parameters that were drawn from a range of plausible parameters based on fits to human behavior. The relative starting time parameter s was taken from the set $\{-0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75\}$, and the standard deviation of the noise, σ, was drawn from $\{0.3, 0.45, 0.6\}$. The weighting parameters for different attributes are always defined relative to one another, therefore, we kept one fixed and varied the other. Specifically, w_t was fixed to 2.2 and w_h was taken from the set, $\{1.9, 2.2, 2.9\}$. Figure C.9 shows the recovered relative start time parameter s for different values of σ and w_h. The plot shows the posterior distribution of the group level mean of the s parameter, and demonstrates that a HstDDM based on our piece-wise constant approximation method can accurately recover known relative starting time parameters.

Figures C.10, C.11 and C.12 show the posterior distributions of the other mean parameters at the group level. As expected based on the recovery analysis for the HaDDM, there is a small overestimation bias for parameters defining the drift scaling terms, i.e. the weight parameters in the stDDM. Thus, once again we checked that the method was able to correctly distinguish between different parameter magnitudes. We calculated the posterior probability distribution of the difference between w_h parameters for different simulations. Figure C.13 confirms that the method is able to distinguish between different magnitude of the weight parameters.

Further evidence to support this claim can be found in analysing the difference in weights $w_h - w_t$. As shown in Webb 2018, the probability of choosing one option over the other in a DDM framework can be approximated by logistic function. Thus, a change in unit of one attribute's weight relative to the other attribute's weight is quantifiable as the exponential of the difference in weights, e.g. $e^{w_h - w_t}$. Figure C.14 shows that the piece-wise constant approximation method is able to correctly recover the difference in weights $w_h - w_t$.

Fitting the HstDDM to groups of agents with heterogeneous model parameters

Next, we tested parameter recovery accuracy for the HstDDM when each agent in the group has a different combination of model parameters. For each subject in the simulated groups, we randomly drew the parameters from gaussian distributions around a specific mean. The means and distributions of the generating parameters are shown in figure C.5b. The model was able to accurately recover the generating parameters at the individual level when each subject had different generating parameters. Figure C.15, C.16 and C.17 show the posterior distributions of the mean parameters at the individual level, for each subject separately. Notably, the parameters are correctly recovered for almost all the subjects in the data set.

Fitting empirical data from human participants

Lastly, we report the results of fitting the HaDDM and HstDDM to two sets of human choice data using our piece-wise constant approximation method. The full details of the experiments that generated these data are listed in the Methods section, and brief summaries of the behavioral paradigms are given in the two subsections below. Overall, simulations from the best-fitting HaDDM and HstDDM parameters can faithfully reproduce the human choice and response time patterns observed in the data they were fit to, indicating that the two models, when fit using the piece-wise constant approximation method, fit human choice data well.

Similar to the above recovery fitting analysis, we assume that participants come from a single population, thus that the hierarchical structure consists of a single group. Based on this, we make the standard assumption that individuals are members of a normally distributed population and assign a normal prior for each individual level parameter.

HaDDM fits to human choice data

We used a real data set of a binary lottery-choice task experiment in which participants had to complete 70 decision trials between a gamble and a sure option, to test the piece-wise constant approximation method for fitting the aDDM. Briefly, in each trial of this task, the sure option offered a certain amount of money with probability = 1, and the gamble option offered a higher amount with some probability p or nothing with probability 1-p. We tracked participants' eye-movements and recorded response times during each trial. Notably, participants were free to look anywhere on the computer screen that displayed the options and to take as long they needed to reach a decision (see Method section for further details).

The aDDM parameters estimated using the piece-wise constant approximation method can accurately recreate the empirical pattern of choices and response times in the lottery data set. In particular, Figure C.18d shows that the simulated and the empirical reaction time distributions conditional to the choice are strikingly similar. This suggests that the estimated

δ, θ, and σ parameters are reasonably accurate.

We ran an additional parameter recovery test after fitting the lottery task data set because the estimated values of certain parameters were outside the range we included in the *a priori* recovery tests above. Specifically, the posterior distribution of the δ parameter at the group level was outside of the range of values we used in the parameter recovery analysis in the previous section. The reason for the lower value of δ in these data is straightforward. The response times (see figure C.18d) in this particular lottery task experiment are slower than those in the food choice data on which we based the parameter ranges in the parameter recovery analysis. Thus, the drift parameters are estimated to be lower in the lottery choices to account for these slower RTs. Therefore, to make sure that lower drift rates do not affect the ability of the model to recover the discount factor parameter, θ, we ran another recovery fitting analysis with the parameter values for δ and σ that we obtained when fitting the model to the empirical data. We generated simulated choice data with the θ parameter set to $\{0.2, 0.4, 0.8\}$ in order to check whether the model was able to accurately recover the three values of θ when δ was 1.2. This additional parameter recovery test indicated that the piece-wise constant approximation method was able to recover these known values of θ even with the relatively lower value of δ (see Figure C.21).

The importance of confirmatory parameter recovery tests

Running parameter recovery tests for the values estimated from a given empirical data set is a good practice. It is especially important when dealing with new data sets. It is often mathematically and computationally challenging to determine *a priori* all the potential combinations of parameter's values for which the model (or the fitting method) is *not* able to properly fit the data or return the true generating parameters. For instance, a low signal to noise ratio certainly affects the ability to accurately recover parameters, but what is "low" exactly? Further, a narrow range of value differences between choice options could affect the ability of a diffusion model to fit the data and derive the generating parameters, but what is a "good" or a "bad" range? Answering these questions with satisfactory precision is not trivial, and therefore, running *a posteriori* parameter recovery checks is best practice. The first step in such a recovery test is to check the ability of the model and fitting procedure to recover relevant known generating parameters. This can be done by simulating the data with the parameters obtained when originally fitting the empirical data, and then re-fitting the model to the simulated choice and response times. However, it is also possible that the recovery fitting fails by always returning the same parameters independently of the generating values. Therefore, success in this first step does not ensure you against this flaw in the model or fitting procedures. Thus, a necessary second step is to make sure that the model and fitting procedures are able to accurately distinguish between different generating parameter values. Specifically, for the second recovery testing step, all the simulating parameters but one are taken from the fitting of the empirical data as in the first step. The one remaining parameter is then varied across a range of plausible values that you want to test. This procedure ensures that the model can recover different magnitudes of that specific parameter of interest under the real data conditions, i.e. all the other parameters equal to the originally estimated values. Ideally, this second testing step would be run for each parameter in the model. At minimum, it should be run for all parameters of interest in the current study.

stDDM fits to human choice data

To test the piece-wise constant approximation method for the stDDM, we used real data from a binary food-choice task experiment in which participants had to complete 50 decision trials between food items. Before performing the food-choice task, participants had to rate all the 180 food items in the choice set for healthiness and tastiness (see Method section for details).

Figure C.19 shows the results of fitting the HstDDM to this food-choice task data set. Noticeably, in figure C.19d the simulated and the empirical reaction time distributions conditional to the choice are compared. As for the HaDDM data fitting, fitted choices and reaction time are remarkably similar to the empirical ones. Further, we performed the confirmatory parameter recovery analysis as described in the section above to make sure the model was able to return the true generating parameters for these specific data (see Figure C.22).

To give a possible estimate of the running time, we also performed some extra fittings varying number of subjects and trials with a standard machine of 4 cpu and 16 GB of ram. We ran 10 subjects with 10 trials each as a time baseline and measured the time increase when doubling number of trials, number of subjects or both. To run 10 subjects and 10 trials, the estimated time was 427 seconds - about 7 minutes. When increasing the number of trials to 20 the time did not doubled, whereas it was estimated to be 660 seconds - about 11 minutes. When increasing the number of subjects to 20 - with still 10 trials each - the time almost doubled to 858 seconds - about 14 minutes. And finally, the time when increasing both trials and subjects to 20 was estimated to 1372 seconds - about 23 minutes.

C.4 Discussion

Here, we presented the explicit derivation of a method to mathematically approximate a pcDDM with a standard DDM given the response time and the interval times in which the drift rate of the pcDDM is not time constant. We provided two practical examples of pcDDMs, the aDDM and the stDDM, and derived their mathematical formulations as standard DDMs. We also used three distinct parameter fitting and recovery analyses to show that this piece-wise constant approximation method was able to accurately estimate the aDDM and the stDDM. The first recovery analysis demonstrated the method for both the aDDM and the stDDM in their hierarchical Bayesian formulations was able to accurately estimate the different magnitudes of parameters at the group and individual levels in simulated data sets. The second recovery analysis showed that the method could accurately recover known heterogeneous individual-level parameters within the hierarchical Bayesian models. Finally, the last analysis tested if the experimental data sets that we used as case studies could be accurately captured by the piece-wise constant approximation method when trying to recover the originally estimated empirical parameters.

Overall, all of the recovery fitting parameter analyses proved to be successful for both the HaDDM and the HstDDM with a high degree of accuracy. Specifically, we could apply this parameter estimation method to the two experimental data sets and show that the HaDDM and the HstDDM reproduced the choice and response time distributions well in both cases. Thus, the piece-wise constant approximation method proved to be a simple, yet powerful, method that allows existing hierarchical Bayesian estimation tools developed for the standard DDM to approximate the time-varying drift rate of pcDDMs. We also showed how fast the model could fit experimental data of standard size and we believe this method can allow researchers not familiar with computational modelling to accurately fitting experimental data set. Nevertheless, we argue here that any model fitting procedure to new data sets, even when methods or models have been tested to be robust and consistent in the field, should be taken with caution, and few preliminary steps should be followed rigorously before making any inference on the fitted parameters. Although we do not explicitly show it in this paper since it is not relevant for the method discussed here, before running any recovery or empirical fitting the predictions of a model should always be tested in the experimental data in order to verify that the model is good approximation of the generating process of interest. Then, a recovery fitting parameter analysis must be implemented to ensure that the model is able to return the known generating parameters' values. This analysis does not necessarily go through all the steps described in the present paper, but as we suggested in the results section a confirmatory parameter recovery test is recommended. This can be done first by simulating the data with the parameters obtained when originally fitting the empirical data and re-fitting the model to the simulated choice and response times. And second, testing a following recovery analysis in which all the simulating parameters but one are taken from the fitting of the empirical data as in the first step. The one remaining parameter is then varied across a range of plausible values that has to be tested. Ideally, this second testing step would be run for each parameter of the model. At minimum, it should be run for all parameters of interest in the current study. Such analysis is good practice to make sure that specific experimental data can be accurately be fitted by the model of interest, and also it is a effective way to test whether the model of interest is identifiable. To conclude, we show that our method is simple, fast and robust in fitting simulated and experimental data with pcDDMs. Given the spread in recent years of such models, we believe that this approach will increase the application of pcDDMs in the field of decision-making.

C.5 Methods

Data sets

Lottery task

Subjects. Thirty-six healthy subjects participated in our experiment. Six of them were excluded from the experiment because they failed to understand the choice task. Subjects received monetary compensation for their participation, were informed about all aspects of the experiment and gave written informed consent. The experiments conformed to the standards of the Declaration of Helsinki and the Human Subjects Committee of the University of Zürich approved the experimental protocol.

Task. The experimental design was divided into two main parts: a decision phase and an elicitation phase. During this decision phase, participants had to make 70 lottery decision trials between a sure option and a gamble. At the beginning of each trial the participant was presented with a monetary endowment (e.g. "You receive 50 CHF"). Ten different starting amounts were used in the experiment (from 10 CHF to 100 CHF with an increment of 10 CHF). To move on to the decision screen, the participants had to fixate for more than 2 seconds the received monetary amount, which was positioned in the centre of the screen. After the fixation period a sure option and a gamble appeared on the screen. The sure option was presented as an amount of money retained from the starting amount (e.g. keep 20 CHF out of a total of 50 CHF). The gamble option always offered the chance to keep all of the starting amount of money with some probability. Seven different probabilities were used in the study, such that the probability of winning in a given trial was either 30%, 40%, 50%, 60%, 70%, 80% or 90%. The expected outcomes, i.e. expected value (EV), of sure and gamble options were always equivalent in each trial. However, the EV varied across trials from 3 to 90 CHF. The sure and the gamble options on the screen were represented through rectangles and pie-charts. Rectangles indicated the amount of money that the subject could keep from the starting amount, pie charts represented probabilities (e.g. the sure option always had a pie-chart completely filled with a color). Before performing the choice task, subjects went through training task to learn the meaning of rectangles and pie-charts.

In the second phase of the experiment, our participants had to perform a certainty equivalence (CE) elicitation task in which they were shown all the 70 gambles that they encountered during the choice task. In each trial, for a total of 70 trials, a gamble and a list of sure amounts of money was shown to the participant. The subject had to make 10 decisions in each trial between the gamble displayed on the left hand-side of the screen and 10 sure amounts of money. The sure amounts of money were displayed in a decreasing order from a maximum value smaller but close to the amount of money that they could win if they choose the gamble, to a minum value close to zero. For instance, a gamble can be described as follows. You first receive an endowment of 100 CHF and you have a probability of 70% of keeping all the 100 CHF (30% of having nothing). In the first line, the participant had to decide between the gamble and a sure amount of 95 CHF. In the second line, she had to decide between the gamble and a sure amount of 85 CHF, and so on until the last line in which she had to choose between the gamble and a sure amount of money of 5 CHF. The CE was calculated as the mean between the two amounts of money where the subject switches from choosing the sure amount to choosing the gamble. Thus, in the example above, if a participant chooses the sure amounts of money for every decision lines until 55 CHF, and she chooses the gamble from 45 CHF on, the CE for this gamble is 50 CHF. We forced participants to only switch from choosing the sure amount to choosing the gamble and only once per trial — including the possibility of never switching, i.e. always choosing the gamble or always choosing the sure amount.

All tasks were programmed in Matlab 2015b (Matworks), using the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007).

Eye Tracking. Before each decision trial, subjects were required to fixate the monetary endowment positioned at the center of the screen for 2 s before the options would appear, ensuring that subjects began every trial fixating on the same location. Subjects' gazes were recorded at 500 Hz with an EyeLink-1000 (http://www.sr-research.com/) eye tracker. Choice trials with no gaze time on any option attribute were excluded from the analysis (17 trials, 0.004% of the pooled data from the 30 subjects).

Food-choice task

Subjects. Ninety-seven healthy subjects participated in our experiment. Eleven of them were excluded from the experiment because they either failed the questionnaire about the choice task - i.e. understanding the task, or because they did not comply with the health agreement. Subjects received monetary compensation for their participation, were informed about all aspects of the experiment and gave written informed consent.

Task. The experimental design was divided into two parts: a rating phase and a decision phase. During the rating phase participants had to rate 180 food items for healthiness and tastiness on a continuous scale from -5 to 5. The food items were shown to the subjects as images on the computer screen with a rating bar below one at the time. After the ratings, subjects had to perform a binary food-choice task consisting of 60 trials in which two food items were presented on the computer screen and they had to choose the food that they would like to receive at the end of the experiment. Before starting the food-choice task, we gave each subject an health agreement instruction form in which the benefits of eating healthy were illustrated and in which subjects were asked to try to make healthy choices during the following food-choice task. In addition, subjects had to express their willingness to try to eat, i.e. choose, healthy check-marking the yes or no option at on the health agreement. Subjects who were not willing to try to eat healthy were subsequently excluded from the analysis.

Model fitting

We fitted and simulated data with R and RJags. We wrote the model in Jags using the dwieners function of the DDM. This function takes as inputs the following parameters: dwieners(α,ndt,β,μ,σ), where α is the boundary separation parameter, ndt the non-decision time, β the bias parameter, d the drift rate parameter and σ the standard deviation of the drift process. Concerning the drift rate μ, we set it to be equal to the function derived in equation 15 and 20 for the aDDM and the stDDM, respectively. Thus, for the aDDM the drift rate function was set to

$$\mu = \frac{\tau_S}{\tau}\delta(V_S - \theta V_G) + \frac{\tau_G}{\tau}\delta(\theta V_S - V_G),$$

where τ is the reaction time, τ_S and τ_G the total fixation time towards the sure option and the gamble respectively, δ is the drift constant parameter, and V_S and V_G the value of the sure option and the gamble, respectively.

For the stDDM, the drift rate function was set to

$$\mu = \begin{cases} \frac{s}{\tau} w_T \mathrm{VD}_T + \frac{\tau - s}{\tau}\left(w_H \mathrm{VD}_H + w_T \mathrm{VD}_T\right) & \text{if } s > 0 \;\wedge\; s < \tau \\ \frac{|s|}{\tau} w_H \mathrm{VD}_H + \frac{\tau - |s|}{\tau}\left(w_H \mathrm{VD}_H + w_T \mathrm{VD}_T\right) & \text{if } s < 0 \;\wedge\; |s| < \tau \\ w_T \mathrm{VD}_T & \text{if } s > 0 \;\wedge\; s > \tau \\ w_H \mathrm{VD}_H & \text{if } s < 0 \;\wedge\; |s| > \tau \end{cases}$$

where w_T is the weight given to the taste attribute, w_H the weight to the health attribute, VD_T and VD_H are the value differences in taste and in health respectively, s is the time at which the health attribute comes into the accumulation process - $s > 0$ means taste is on from the beginning and health comes in at time s, $s < 0$ means that the health attribute is on the accumulation process from the beginning and taste comes in at time $|s|$.

For both models, we fixed the boundary separation parameter $\beta = 2$, and estimated the standard deviation of the noise σ and the non-decision time ndt. Specifically for the aDDM, we estimated the drift rate constant δ and the discount factor θ, whereas the total fixation times τ_S and τ_G were given as an input to the model. Concerning the stDDM, we estimated the weights parameters w_T and w_H, and the relative starting time parameter s. All the detailed code of the implementation and the data can be found at https://github.com/galombardi/method_HtSSM_aDDM.

C.6 Figures

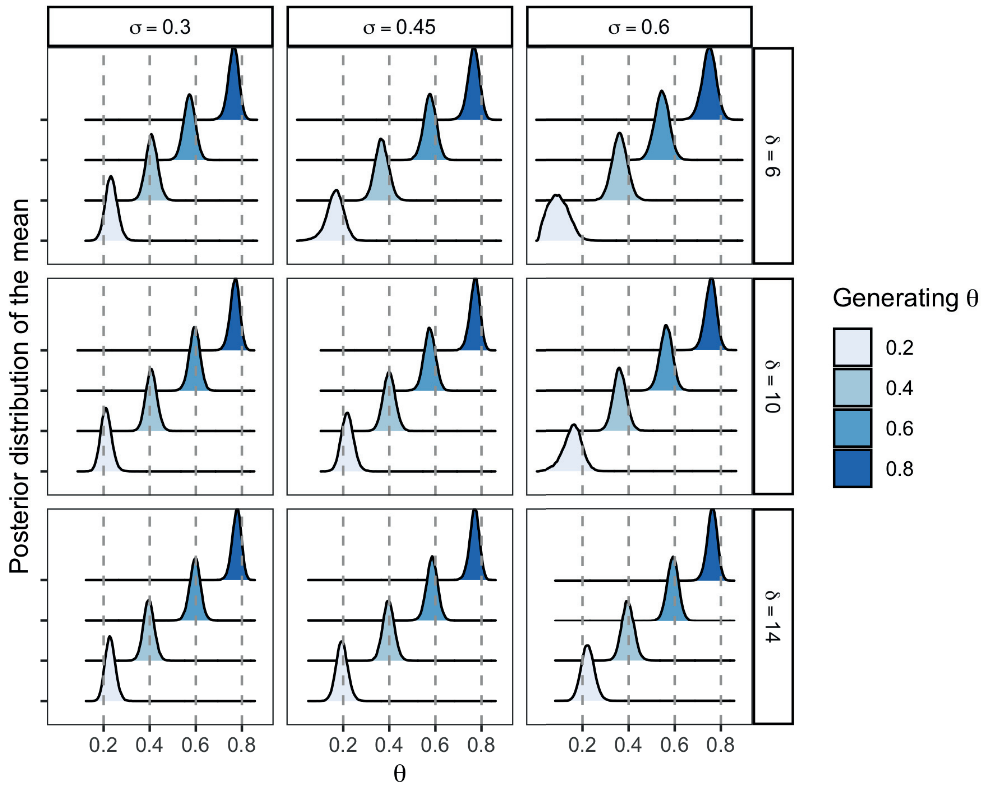


Figure C.1: Group level posterior probability distributions of the mean θ parameters from the HaDDM fits of simulations varying the δ and the σ parameters. The dashed grey lines indicate the input generating parameters for the simulations.

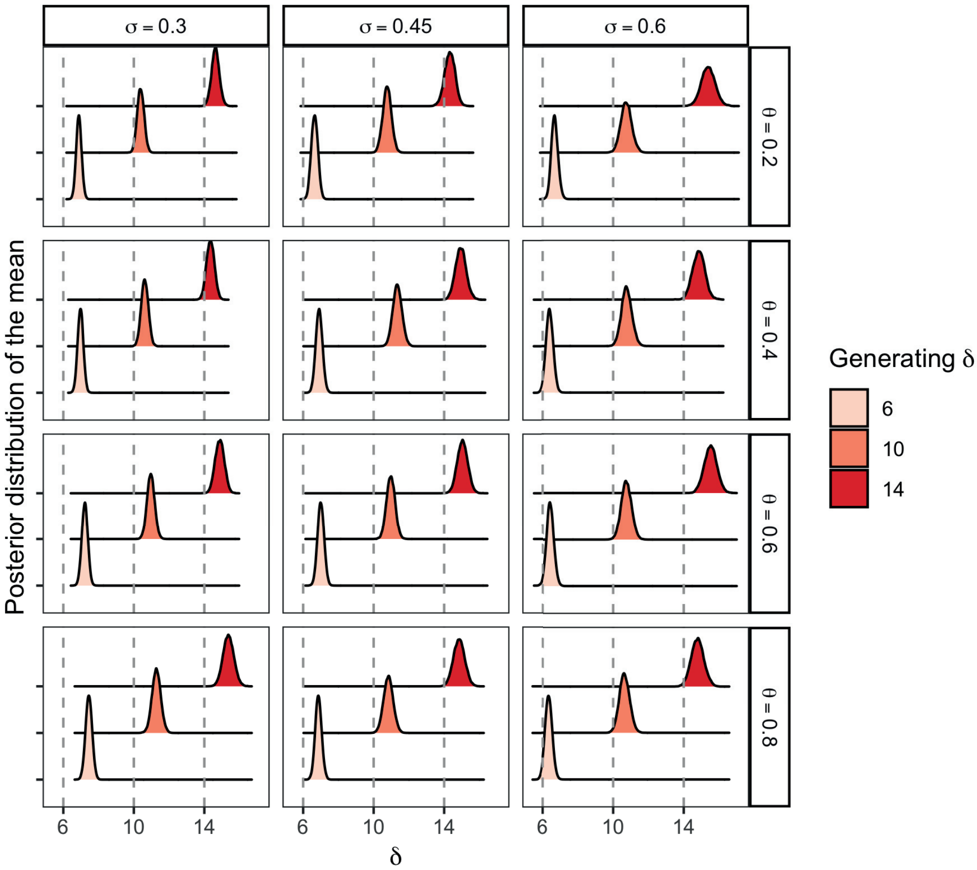


Figure C.2: Group level posterior probability distributions of the mean δ parameters from the HaDDM fits of simulations varying the θ and the σ parameters. The dashed grey lines indicate the input generating parameters for the simulations.

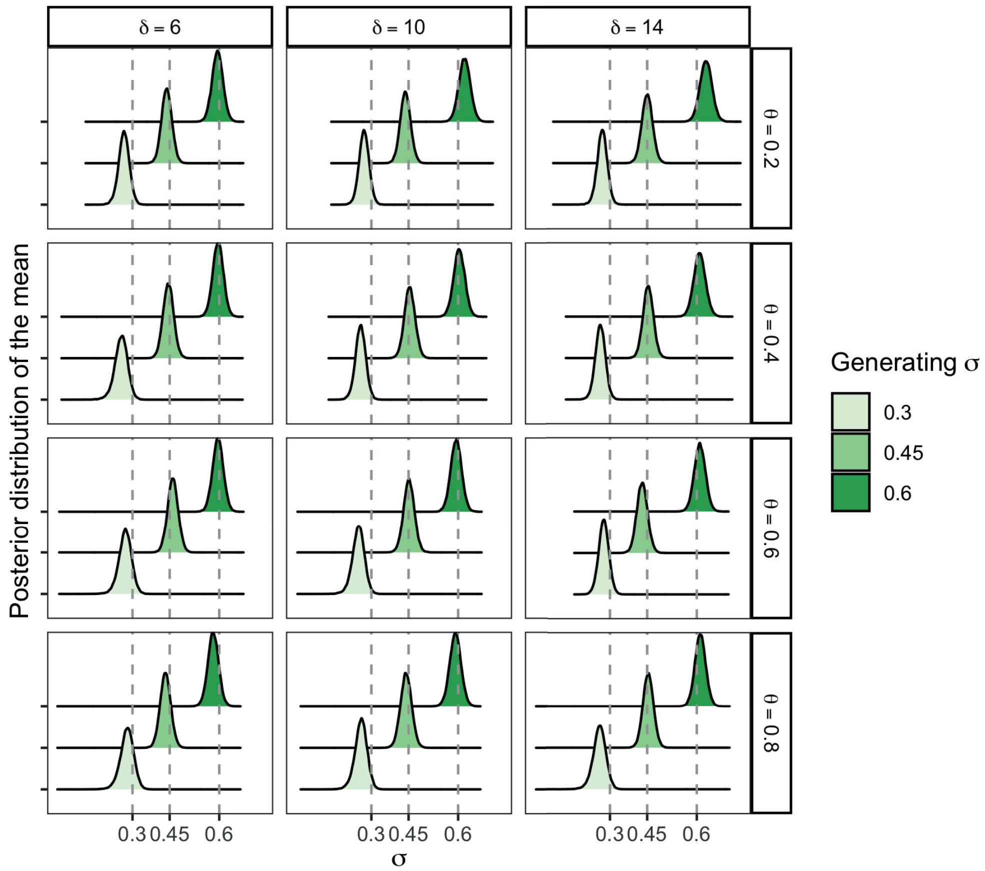


Figure C.3: Group level posterior probability distributions of the mean σ parameters from the HaDDM fits of simulations varying the δ and the θ parameters. The dashed grey lines indicate the input generating parameters for the simulations.

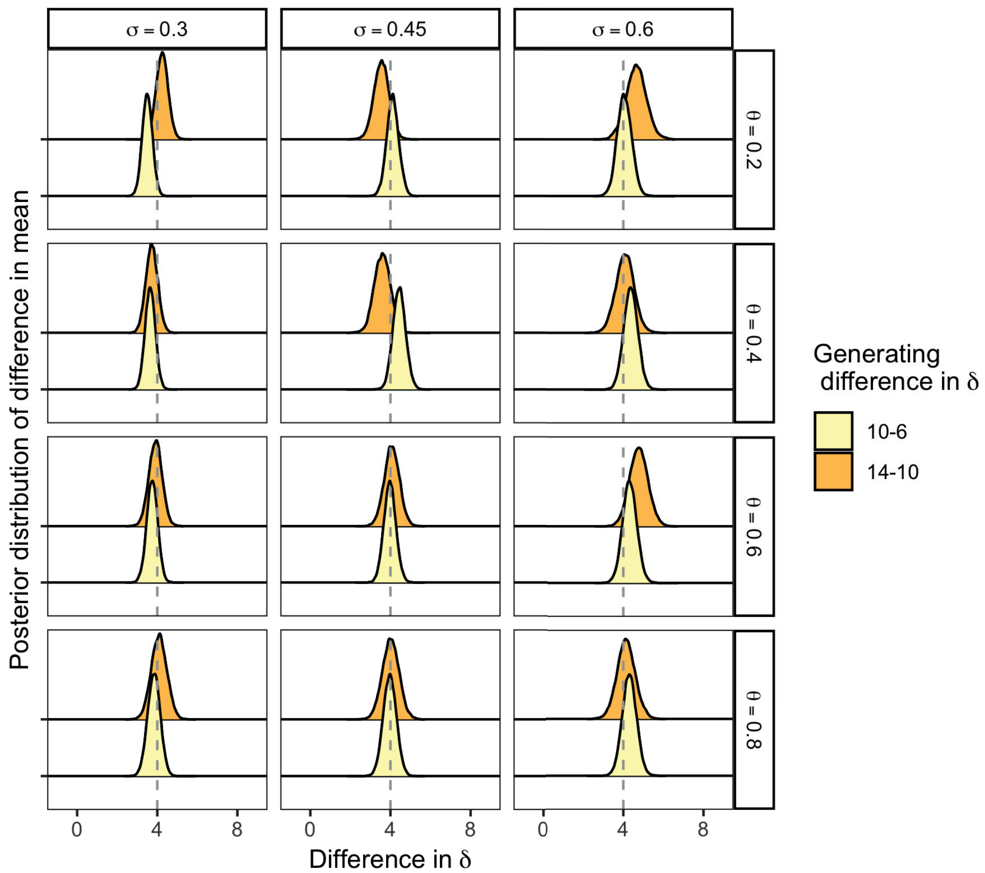
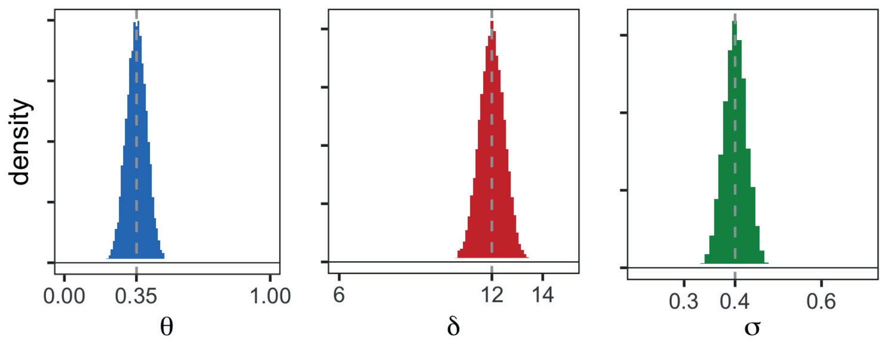
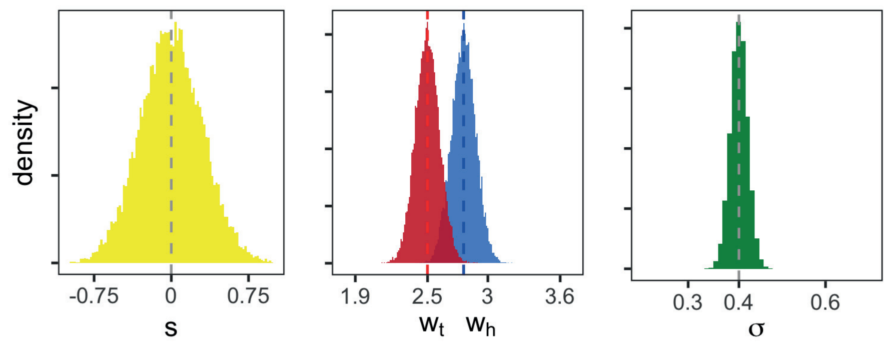


Figure C.4: Group level posterior probability distributions of the difference in mean between δ parameters from HaDDM fits of separate simulations varying the θ and the σ parameters. The dashed grey lines indicate the input generating parameter difference between simulations.



(a)



(b)

Figure C.5: Distributions for the generating parameters in the second step of the recovery analysis procedure. **a)** Generating gaussian distributions $N(\mu,\sigma^2)$ of the aDDM parameters. $\theta \sim N(0.35, 0.0025)$, $\delta \sim N(12, 0.25)$, $\sigma \sim N(0.41, 0.00064)$. **b)** Generating gaussian distributions $N(\mu,\sigma^2)$ of the stDDM parameters. $s \sim N(0, 0.09)$, $w_t \sim N(2.5, 0.01)$,$w_h \sim N(2.8, 0.01)$, $\sigma \sim N(0.41, 0.0006)$.

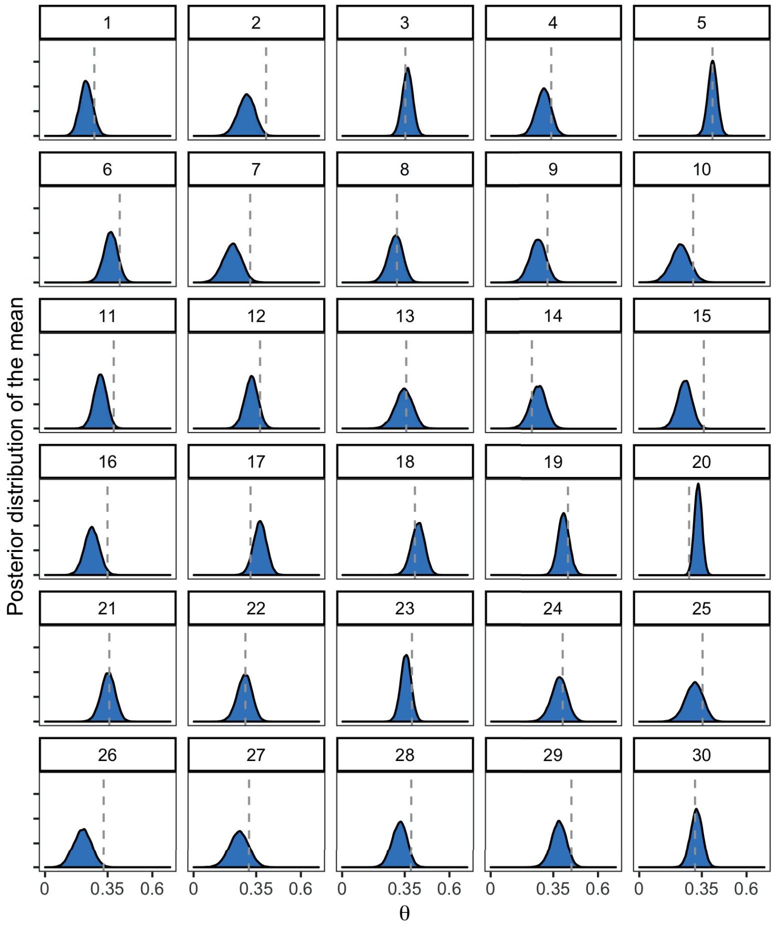


Figure C.6: Individual level posterior probability distributions of the mean θ parameters from the HaDDM fits for each subject in the data set. The dashed grey lines indicate the input generating parameters used for the simulations.

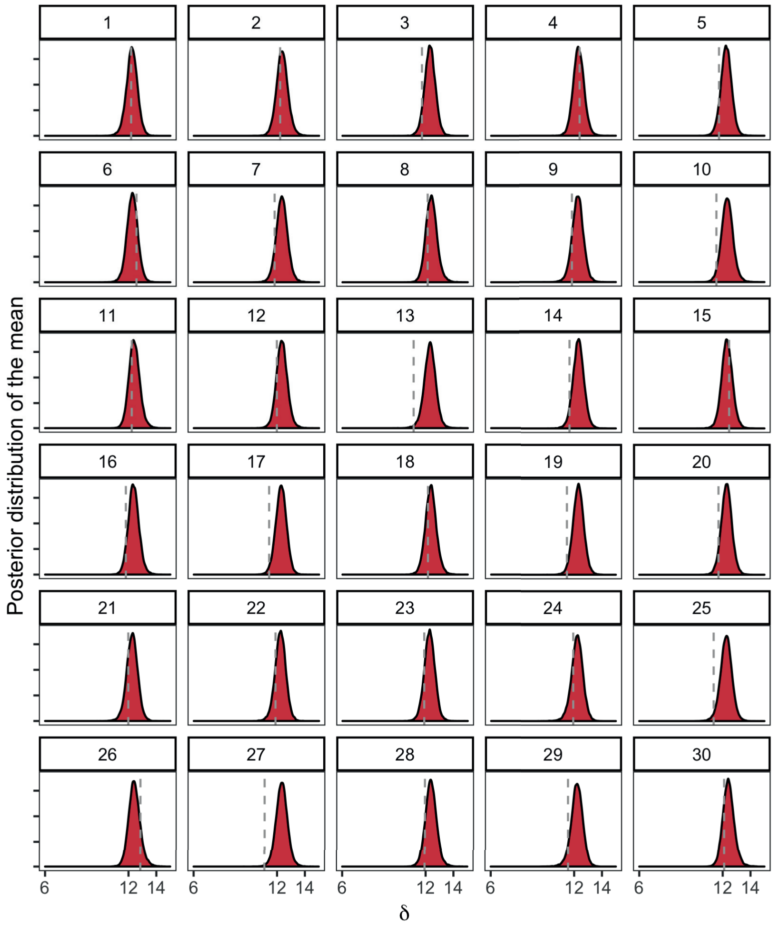


Figure C.7: Individual level posterior probability distributions of the mean δ parameters from the HaDDM fits for each subject in the data set. The dashed grey lines indicate the input generating parameters used for the simulations.

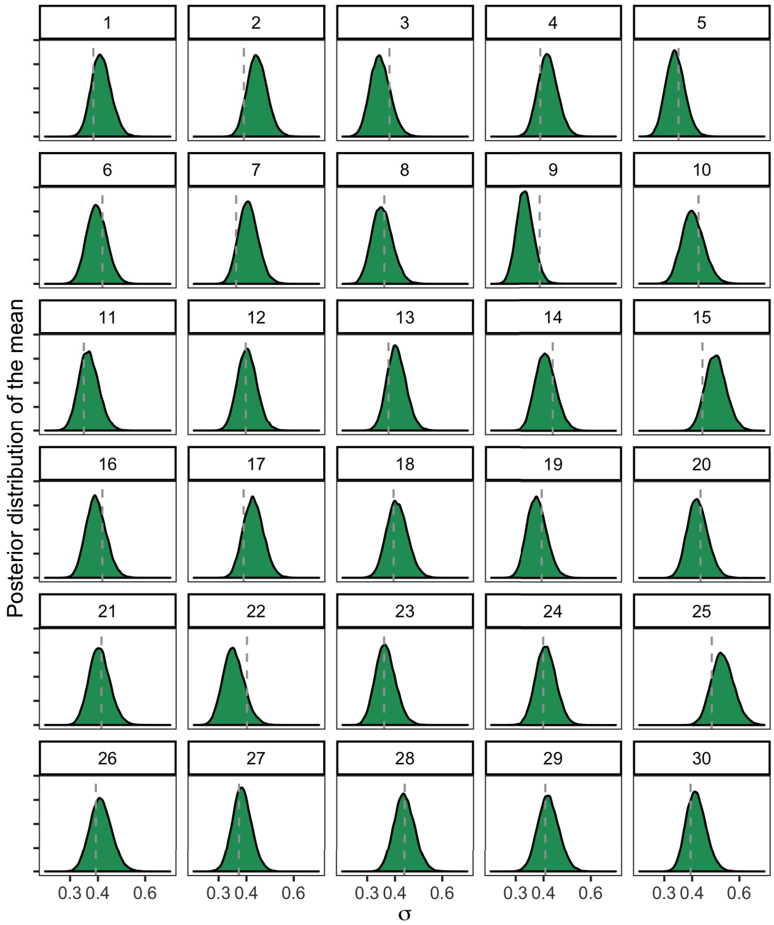


Figure C.8: Individual level posterior probability distributions of the mean σ parameters from the HaDDM fits for each subject in the data set. The dashed grey lines indicate the input generating parameters used for the simulations.

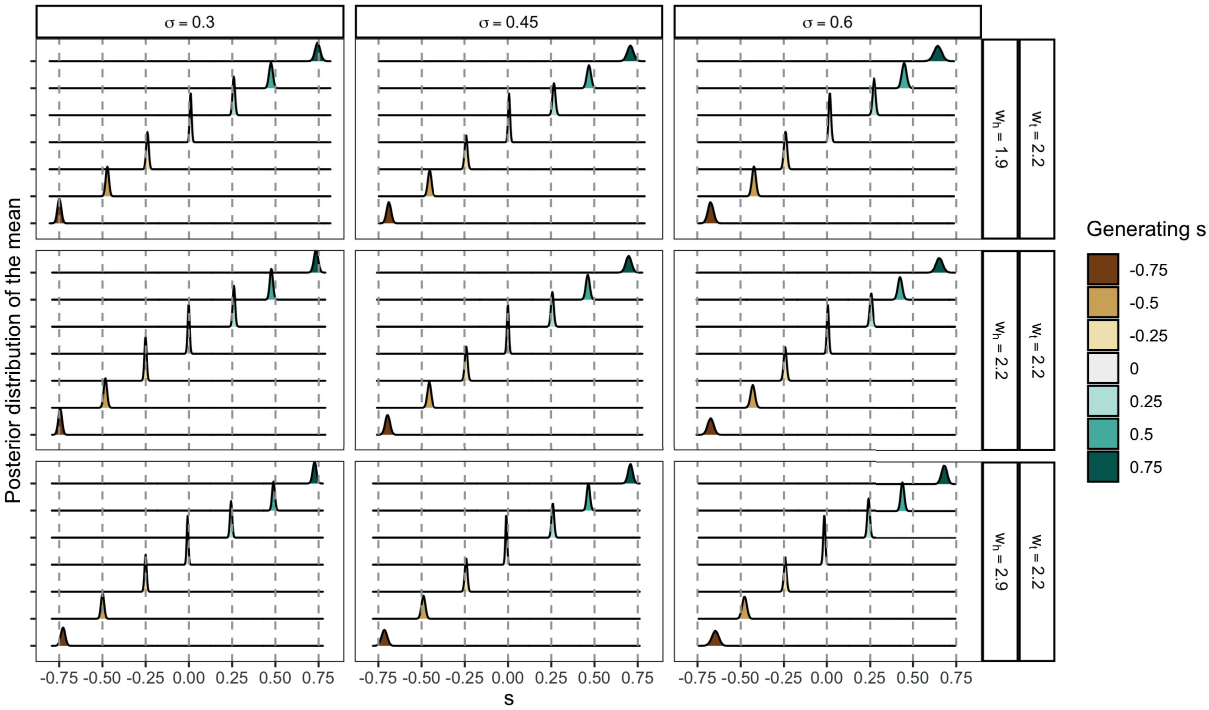


Figure C.9: Group level posterior probability distributions of the mean s parameters from the HstDDM fits of simulations varying the weights and the σ parameters. The dashed grey lines indicate the input generating parameters for the simulations.

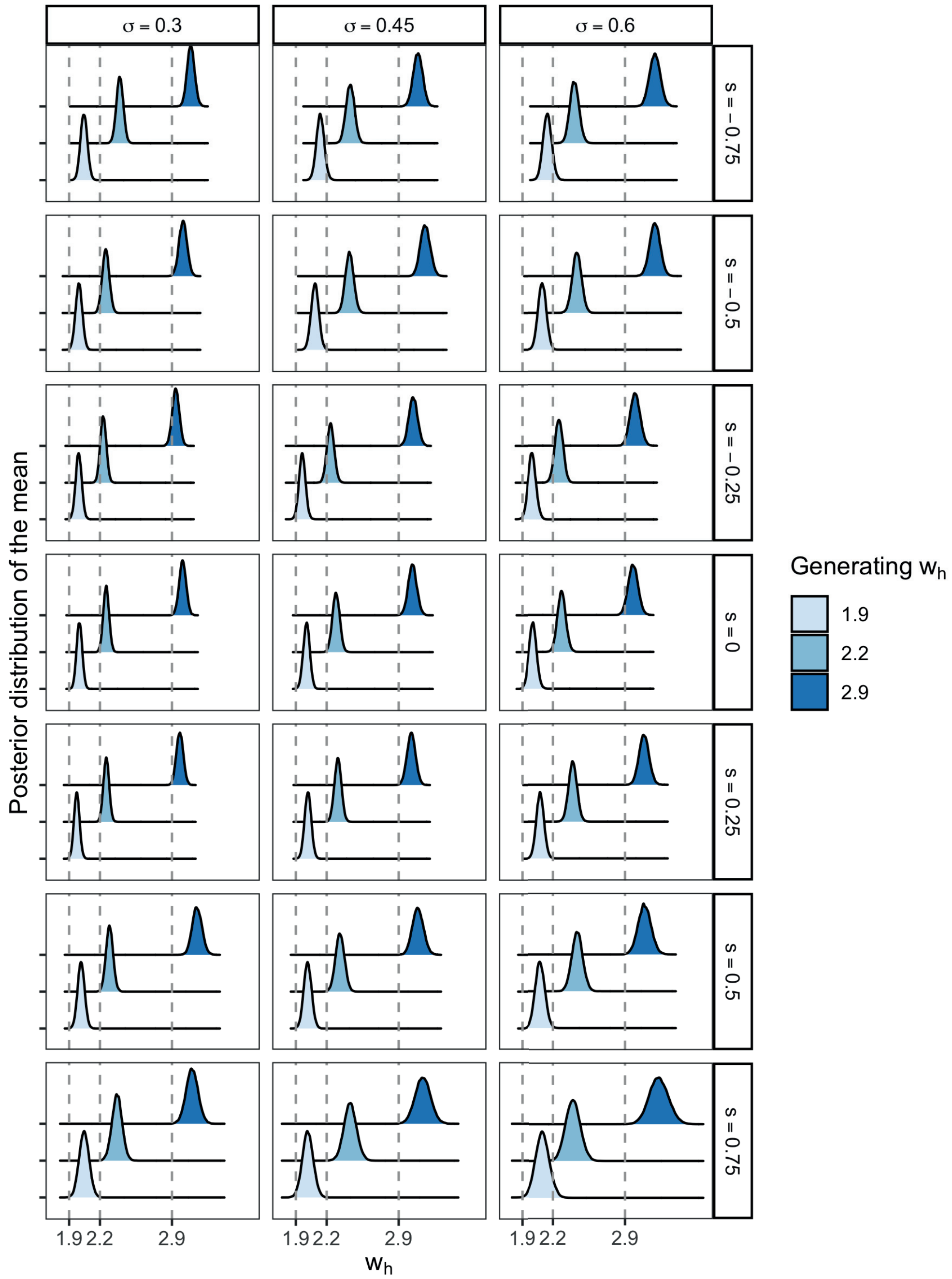


Figure C.10: Group level posterior probability distributions of the mean w_h parameters from the HstDDM fits of simulations varying the s and the σ parameters. The dashed grey lines indicate the input generating parameters for the simulations.

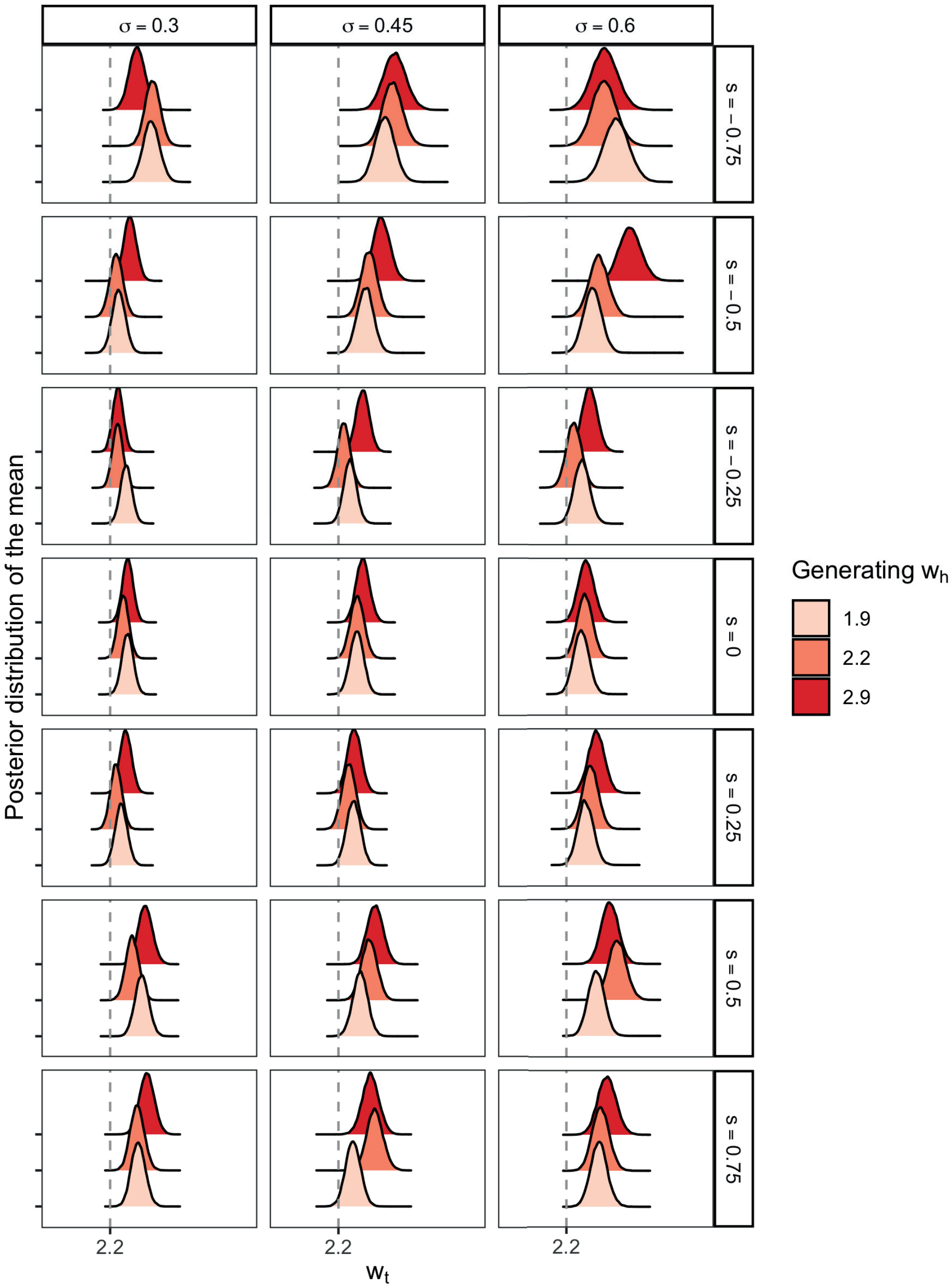


Figure C.11: Group level posterior probability distributions of the mean w_t parameters from the HstDDM fits of simulations varying the s and the σ parameters. The dashed grey lines indicate the input generating parameters for the simulations.

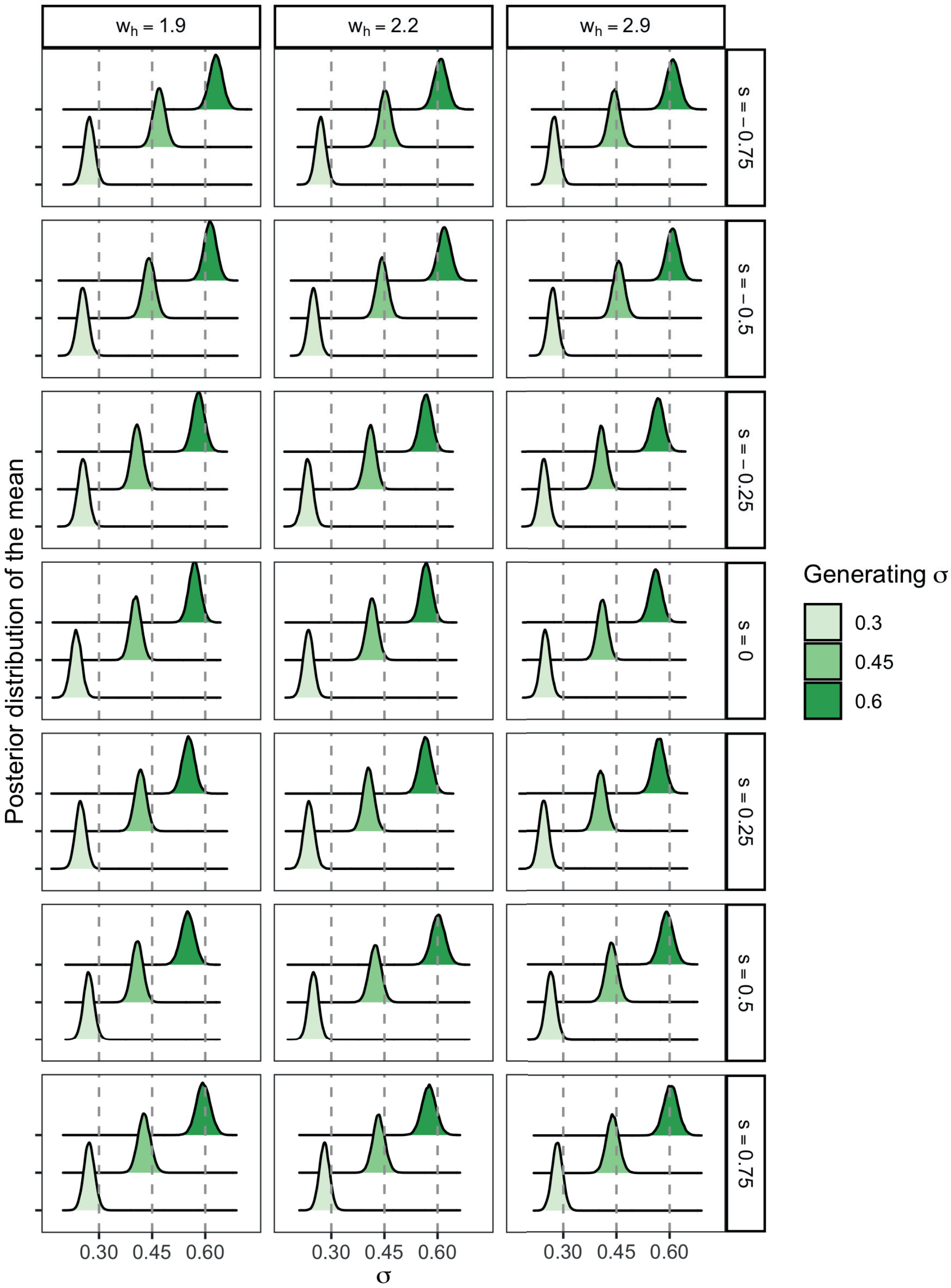


Figure C.12: Group level posterior probability distributions of the mean σ parameters from the HstDDM fits of simulations varying the weights and the s parameters. The dashed grey lines indicate the input generating parameters for the simulations.

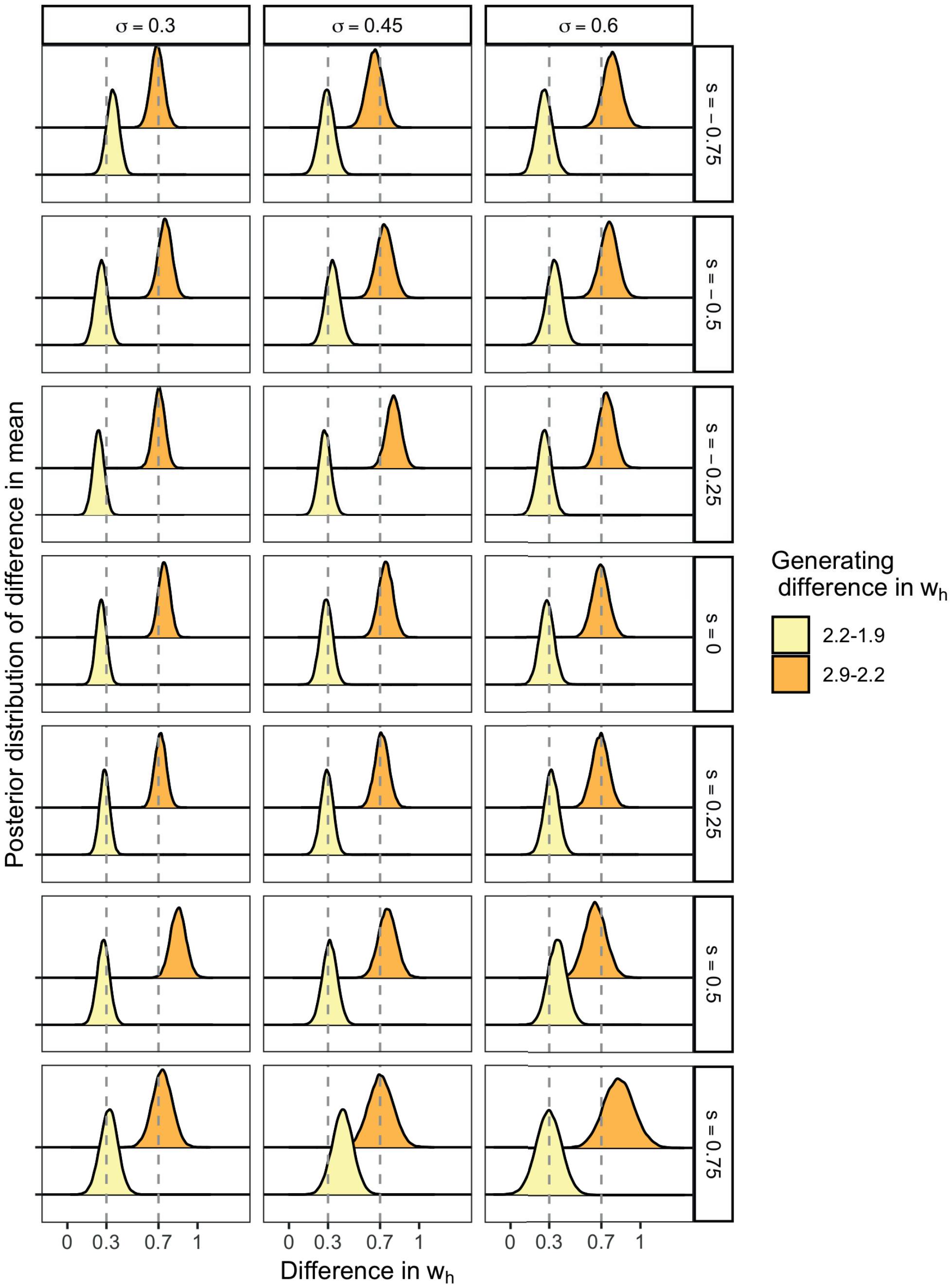


Figure C.13: Group level posterior probability distributions of the difference in mean between w_h parameters from HstDDM fits of separate simulations varying the s and the σ parameters. The dashed grey lines indicate the input generating parameter difference between simulations.

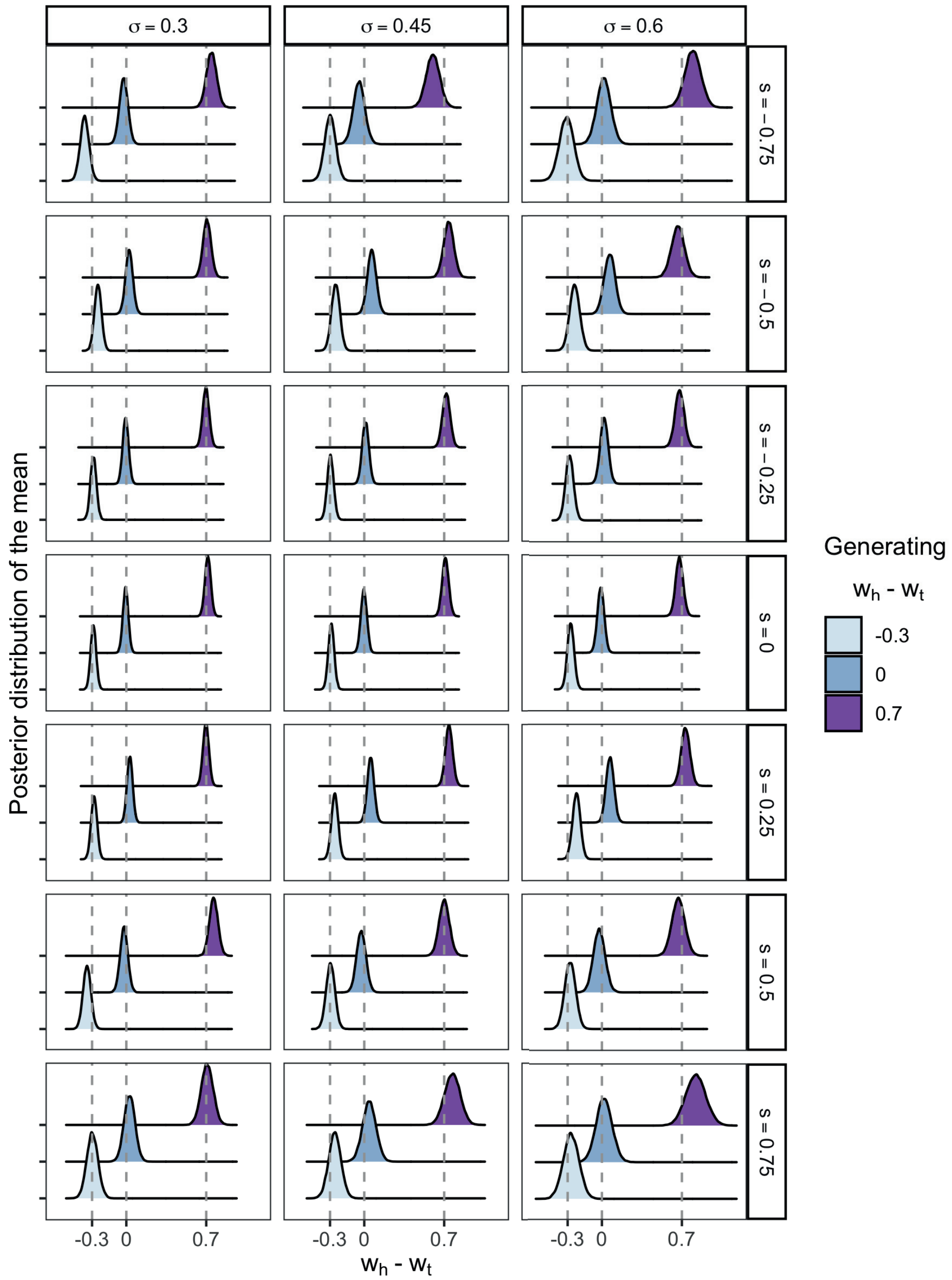


Figure C.14: Group level posterior probability distributions of the difference in mean between w_h and w_t from HstDDM fits of simulations varying the s and the σ parameters. The dashed grey lines indicate the input generating parameter differences between the weights.

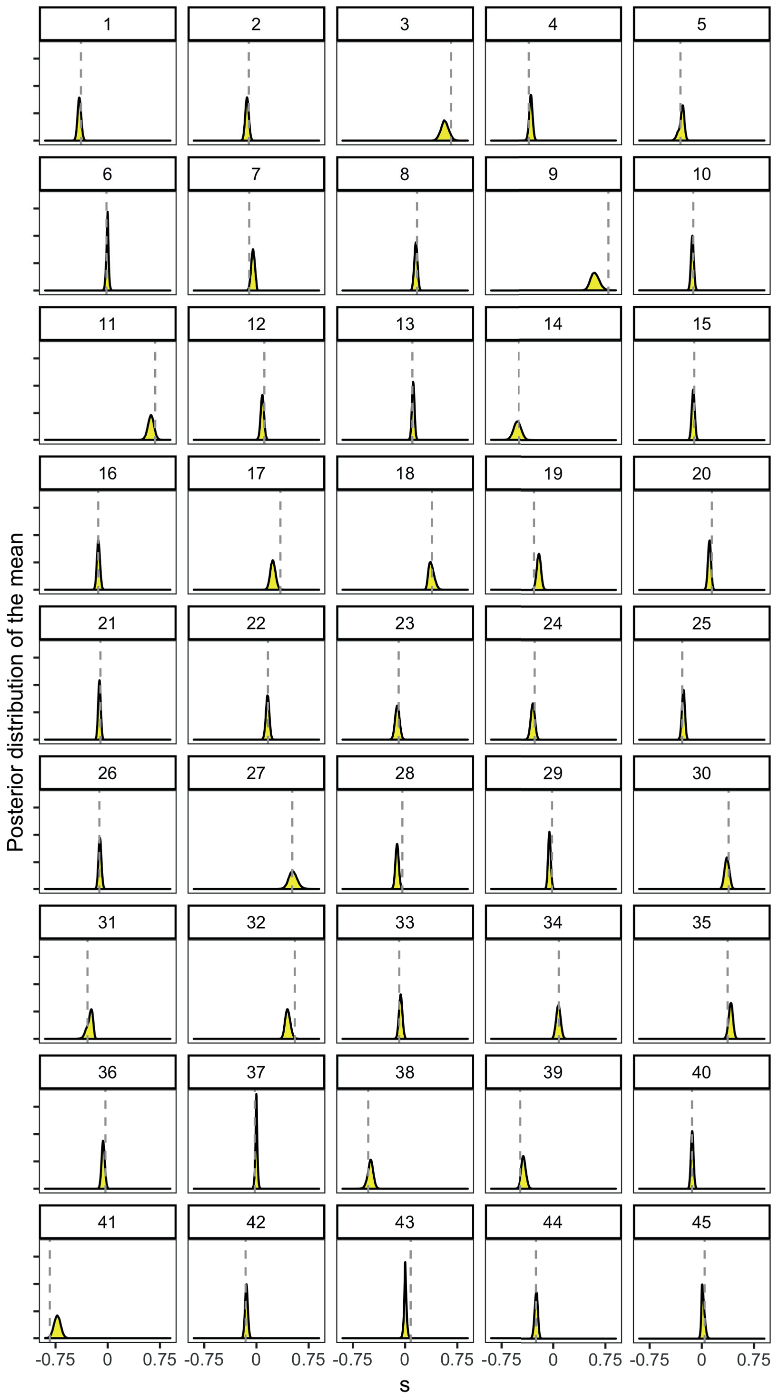


Figure C.15: Individual level posterior probability distributions of the mean s parameters from the HaDDM fits for each subject in the data set. The dashed grey lines indicate the input generating parameters used for the simulations.

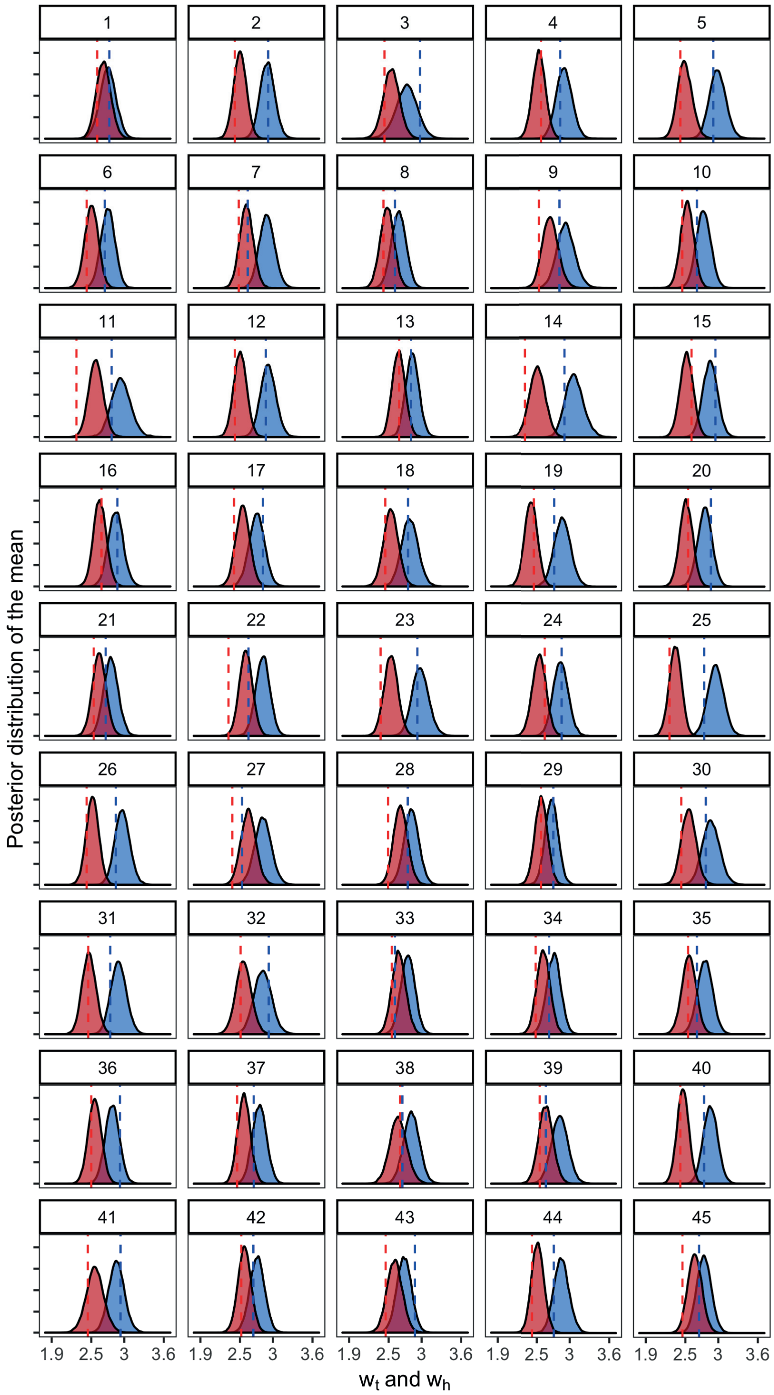


Figure C.16: Individual level posterior probability distributions of the mean w_h,(in blue) and w_t (in red) parameters from the HstDDM fits for each subject in the data set. The dashed lines indicate the input generating parameters used for the simulations.

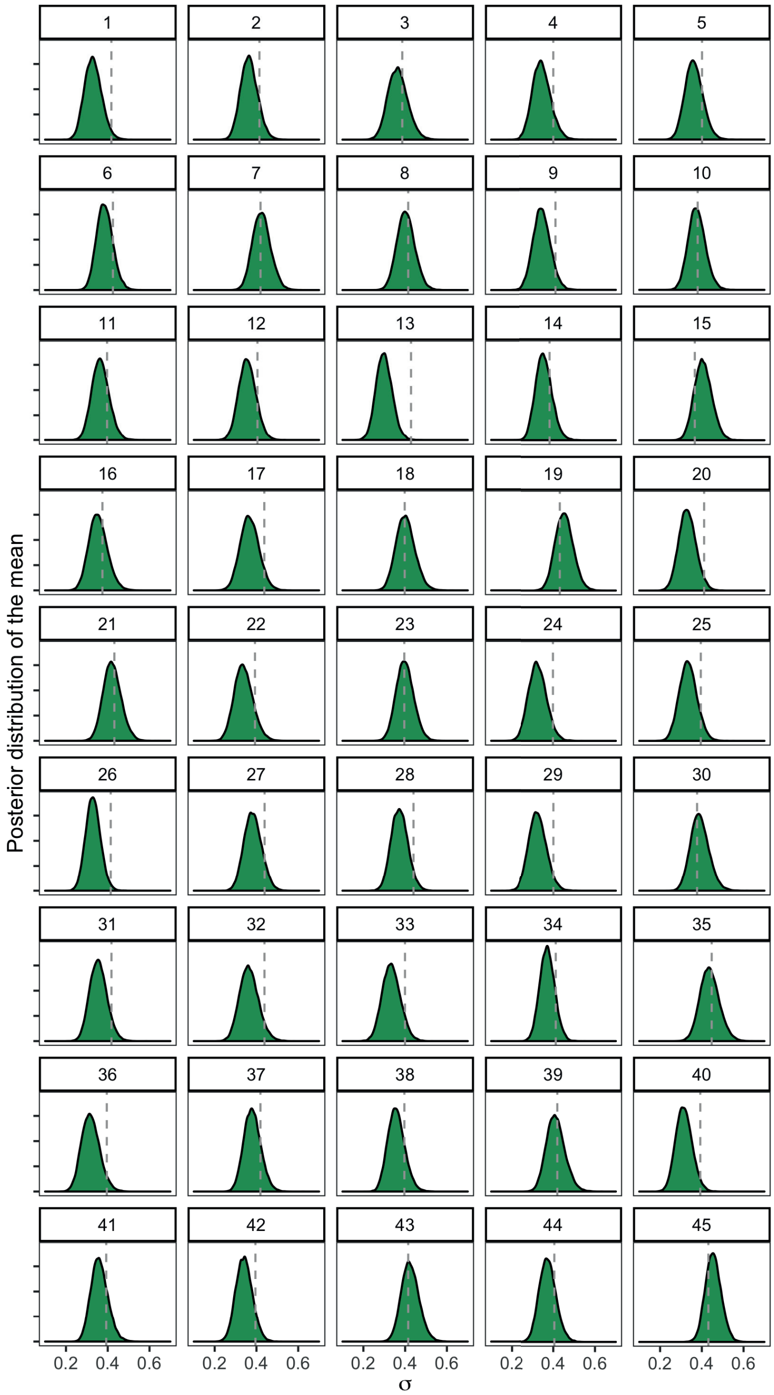


Figure C.17: Individual level posterior probability distributions of the mean σ parameters from the HstDDM fits for each subject in the data set. The dashed grey lines indicate the input generating parameters used for the simulations.

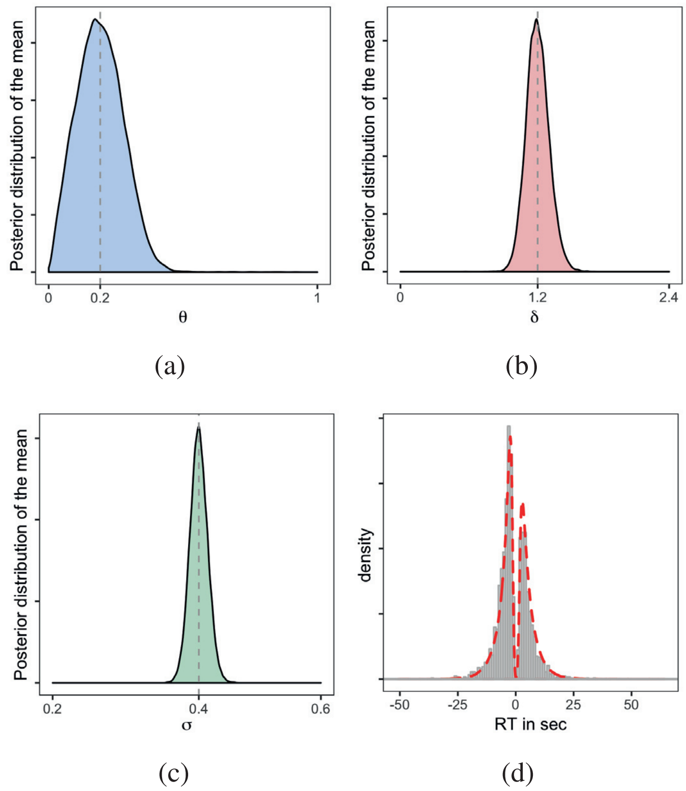
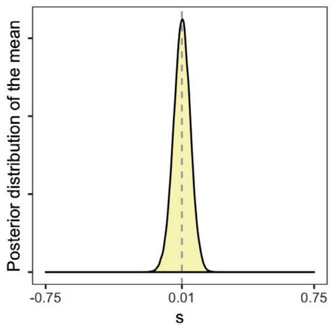
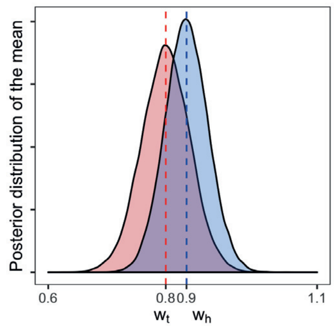


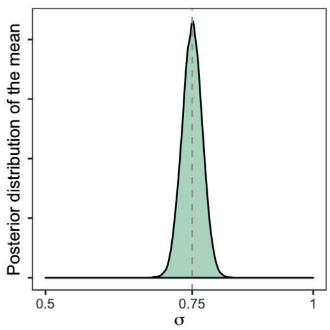
Figure C.18: Model fitting analysis of experimental data with the HaDDM. **a)** Posterior probability distribution of the θ mean at the group level. **b)** Posterior probability distribution of the δ mean at the group level. **c)** Posterior probability distribution of the σ mean at the group level. **d)** Goodness of fit plot of the reaction time for choosing the sure option (negative values) or the gamble (positive values). The red dashed line is the simulated reaction time with the fitted parameters and the grey histogram indicates the empirical reaction time.



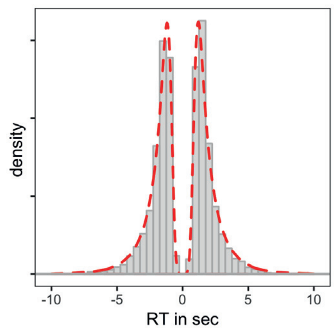
(a)



(b)

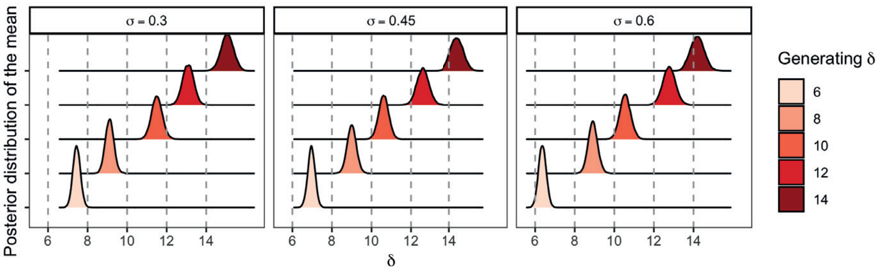


(c)

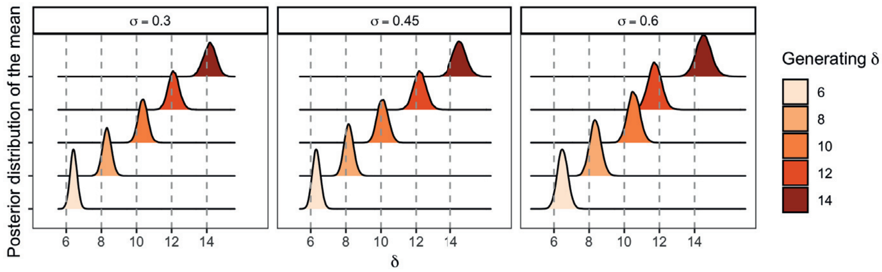


(d)

Figure C.19: Model fitting analysis of experimental data with the HaDDM. **a)** Posterior probability distribution of the s mean at the group level. **b)** Posterior probability distributions of the w_h and w_t means at the group level. **c)** Posterior probability distribution of the σ mean at the group level. **d)** Goodness of fit plot of the reaction time for choosing the right option (negative values) or the left option (positive values). The red dashed line is the simulated reaction time with the fitted parameters and the grey histogram indicates the empirical reaction time.



(a)



(b)

Figure C.20: Group level posterior probability distributions of the δ mean from HDDM fits of separate simulations. The dashed grey lines indicate the input generating parameter. **a)** Simulations were performed with the discrete bounded accumulation series of the DDM - i.e. Euler method. **b)** Simulations were performed with the continuous SDE version of the DDM.

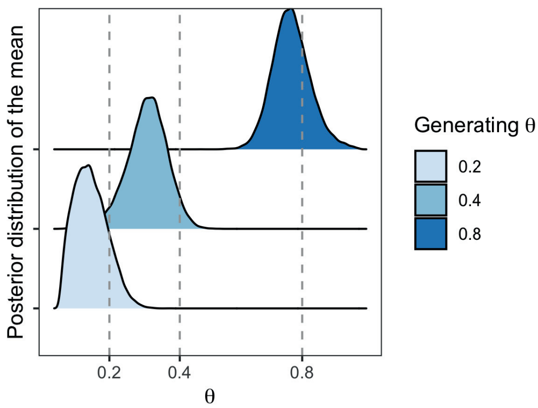


Figure C.21: Group level posterior probability distributions of the mean θ parameters from the HaDDM fits of simulations with δ and the σ parameter values from the model fitting of the experimental data. The dashed grey lines indicate the input generating parameters for the simulations.

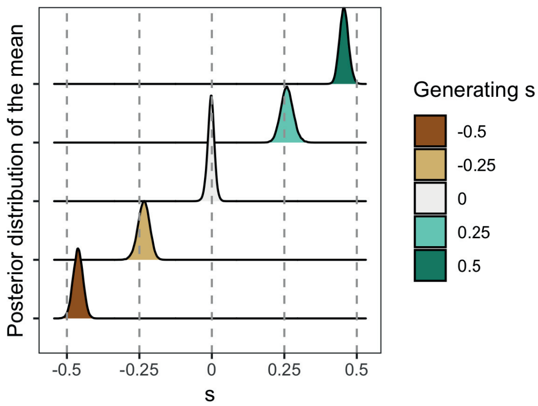


Figure C.22: Group level posterior probability distributions of the mean s parameters from the HstDDM fits of simulations with w_h, w_t and the σ parameter values from the model fitting of the experimental data. The dashed grey lines indicate the input generating parameters for the simulations.

Appendix D:

Thanks to
Todd Hare
Ernst Fehr
Mum
Dad
Aunty Fiò
Robi
Bea
Fa
Elena
Marius
Andy
Rafa
Karl
Carol
Alex
Silvia
Adi
Lisa
Camilla
Elaine
Sebastian
Dan
Camil B

Table D.1: List of people who made this dissertation possible. By any means, the order of appearance is by importance, however, I decided to position first the names of my supervisors who actively worked on the projects and spent their time, knowledge and patience to make this possible. Note, I have never been good with words - which is ironic considering the more than 200 pages of this dissertation, thus, I hope the people on this list are aware of their contribution without me formally expressing it.

Appendix E:

Curriculum Vitae

Personal details

	Gaia Lombardi
Date of birth	27.04.1988

Education

September 20l5 - July 2020	Doctoral program in Neuroeconomics University of Zurich, Department of Economics
February 20ll - December 2013	Master of science in Engineering Mathematics Politecnico di Milano
October 2007 - February 2011	Bachelor of science in Engineering Mathematics Politecnico di Milano

---

Appendix F:

Bibliographic citation

Simple method to fit hierarchical Bayesian piecewise time-constant drift diffusion models. Gaia Lombardi and Todd Hare.

A version of this paper has been submitted to Neurons, Behavior, Data Analysis and Theory (NBDT) .